

Access DB# 64900

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner #: _____ Date: _____
 Art Unit: _____ Phone Number 30 _____ Serial Number: _____
 Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

Jan Delaval
 Reference Librarian
 Biotechnology & Chemical Library
 CM11 1E07 - 703-308-4488
 jan.delaval@uspto.gov

STAFF USE ONLY

	Type of Search	Vendors and cost where applicable
Searcher: <u>Jan</u>	NA Sequence (#) _____	STN _____
Searcher Phone #: <u>4448</u>	AA Sequence (#) <u>✓</u>	Dialog _____
Searcher Location: _____	Structure (#) _____	Questel/Orbit _____
Date Searcher Picked Up: <u>4/22/02</u>	Bibliographic _____	Dr.Link _____
Date Completed: <u>4/22/02</u>	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time: _____	Fulltext _____	Sequence Systems <u>✓</u>
Clerical Prep Time: <u>10</u>	Patent Family _____	WWW/Internet _____
Online Time: <u>4:00</u>	Other _____	Other (specify) _____

This Page Blank (asp.c)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 22, 2002, 10:06:15 ; Search time 14.69 seconds

(without alignments)
2136.414 Million cell updates/sec

Title: US-09-617-923-2

Perfect score: 2099
Sequence: 1 MAEPLRGSRGSRGSGARR.....GKEPLADLGESGLIEVWG 412

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	154.5	7.4	1217	2	hypothetical prote
2	145.5	6.9	1072	1	neurofilament trip
3	140	6.7	1634	2	hypothetical prote
4	137.5	6.6	854	2	neurofilament trip
5	134.5	6.4	587	1	65k early nonstruc
6	133	6.3	971	2	hypothetical prote
7	132.5	6.3	792	2	hypothetical prote
8	132	6.3	586	1	65k early nonstruc
9	128.5	6.1	635	2	nucleolin homolog
10	128.5	6.1	734	2	nucleolin homolog
11	128	6.1	303	1	gene 8 protein - p
12	127	6.1	1365	2	suppressor two of
13	126	6.0	856	2	hypothetical prote
14	124.5	5.9	1684	2	gravin - human
15	124.5	5.9	1791	2	hypothetical prote
16	122	5.8	1082	2	hypothetical prote
17	120	5.7	669	2	hepatoma-derived g
18	119	5.7	450	2	neurofilament trip
19	119	5.7	765	2	gene posterior sex
20	119	5.7	1603	2	hypothetical prote
21	119	5.7	1611	2	hypothetical prote
22	118.5	5.6	508	2	hypothetical prote
23	118	5.6	577	2	hypothetical prote
24	117.5	5.6	1110	2	NF-180 - sea lamp
25	117.5	5.6	2187	2	nascent polypeptid
26	117	5.6	553	2	hypothetical prote
27	116.5	5.6	714	2	hypothetical prote
28	116.5	5.6	955	2	male-specific leth
29	116.5	5.6	974	2	iodestar maternal-

30	116	5.5	672	2	H86169	hypothetical prote
31	116	5.5	1043	2	A56037	DNA-binding protei
32	115.5	5.5	872	2	S62061	SCDS protein - yea
33	115.5	5.5	1280	2	T00365	hypothetical prote
34	115	5.5	382	2	T14336	RAD3 protein, iso
35	114	5.4	891	2	G84693	hypothetical prote
36	114	5.4	1274	2	T16251	hypothetical prote
37	113.5	5.4	2361	2	T25752	hypothetical prote
38	113	5.4	611	2	T06458	nucleolin homolog
39	113	5.4	776	2	T20738	hypothetical prote
40	112.5	5.4	95	2	UC4760	SM73 protein - hum
41	112.5	5.4	675	2	T03744	myod protein inhib
42	112.5	5.4	2109	2	T33247	hypothetical prote
43	112.5	5.4	3942	2	T42730	Bassoon protein -
44	112	5.3	480	2	JC7552	Shb-like adapter p
45	112	5.3	900	2	C96842	hypothetical prote

ALIGNMENTS

RESULT 1
T00270
hypothetical protein KIAA0596 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #extl_change 21-Jul-2000
C:Accession: T00270
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, DNA Res. 5, 31-39, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The comp
A:Reference number: 214086; MUID:98290545
A:Accession: T00270
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-1217 <NAG>
A:Cross-references: EMBL:AB011168; NID:g3043715; PIDN:BAA25522.1; PID:g3043716
A:Experimental source: brain
C:Genetics:
A:Note: KIAA0596

Query Match	7.4%	Score 154.5;	DB 2;	Length 1217;
Best local similarity	23.5%	Pred. No. 0.086;		
Matches 105;	Conservative 48;	Mismatches 164;	Indels 129;	Gaps 21;
QY	6	RGGRPSRGRGRARRARGARCRPRARQSPARLIPOTVLVDVSDSEVLEVADPEVVP	65	
DB	455	RQGRGRQGRSPGRASG-----PNRHQAPSMLSPPA---LSSDSKDEGDEGTEELP	506	
QY	66	-----VARLPAPAKPPRODSDSDEGAEE-----GPAGAPRTLVRRRRRL	106	
DB	507	ALPVLAKSTIKKALASVPSPALPRLPSLHMEKRAQESVGLDPAPAPANPGC---RRGRWV	563	
QY	107	DPEGAPVVPYSGKVOSSINLIPDNSSLKLCPS--EP-EDPADLTNSGSSPEDDALPS	163	
DB	564	QPG-----VELSVSMIDLRLQETLAPSLDDPSQSLATLIPSGRRKKGQALFET	612	
QY	164	G-----SPRRKKLRKCC-----EKKE-----KMEFPQODISPLPQPS---SRNK	201	
DB	613	SLTSQNEKPPRQASQPCSYPIIRLLSQEGVADLEPAPLEIDIVYPEPSDNPMTDT	672	
QY	202	SRKHTALQK-----LREYNKRLOD-----LRSLSPKQHOAPALOSTDEVVU	245	
DB	673	SEQVQAPARAGTIGRVPYPSRSEKSPSACSVYSSSLSPER--PREDESETEPLS	730	
QY	246	VEGPVLPOSSRLFTLTKICRADLVRLVRKSEPLQNVVDMANHLGVSPNRILLFGESE	305	
DB	731	VDD-----ISSDLEPABDEDEEEEGGMP-----YGLQE	762	
QY	306	LSPTAP-----SPLKIGVADIIDICVYLASSSEATFESQELRLVQGRKKQMLE	355	
DB	763	GSP-OTPDQOFLKQHFETLASGAPGAP-VQVPERSESRSISSRPLLVQVQTRPLREP--	818	

356 ISLSPDKLVMSHYEAMGLSGHK 381
 || | ||| : || :
 Db 819 ---SPSSSLALMSRPAQVPAQSGEQ 841

 RESULT 2
 A37221
 neurofilament triplet H protein - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Dec-1999
 C:Accession: A37221; A25649; A30796; A32757; B25649
 R:Chin, S.S.M.; Llem, R.K.H.
 J. Neurosci. 10, 3714-3726, 1990
 A>Title: Transfected rat high-molecular-weight neurofilament (NF-H) coassembles with vimentin
 A:Reference number: A37221; MUID:91038277
 A:Accession: A37221
 A>Status: Preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-1072 <CHI>
 A:Cross-references: GB:A0031879; NID:92642597; PID:AAB87068.1; PID:92642598
 R:Robinson, P.A.; Wion, D.; Anderson, B.H.
 FEBS Lett. 209, 203-205, 1986
 A>Title: Isolation of a cDNA for the rat heavy neurofilament polypeptide (NF-H).
 A:Reference number: A25649; MUID:87080760
 A:Accession: A25649
 A:Molecule type: mRNA
 A:Residues: 230-318:472-542 <ROB>
 A:Cross-references: GB:M37227
 R:Daugherty, A.; Pham-Dinh, D.; Rousset, C.; Felix, J.M.; Nussbaum, J.L.; Jolles, P.
 Biochem. Biophys. Res. Commun. 154, 1099-1106, 1988
 A>Title: The large neurofilament subunit (NF-H) of the rat: cDNA cloning and in situ detection
 A:Reference number: A30796; MUID:88309090
 A:Accession: A30796
 A:Molecule type: mRNA
 A:Residues: 266-421, 'T', 423-427, 'T', 429-542, 'V', 556-566, 'E', 568-613, 'A', 615-725, 'S', 727-733, 'S'
 A:Cross-references: GB:M21964; NID:9205685; PID:AAA41695.1; PID:9205686
 R:Lieberburg, I.; Splinter, N.; Snyder, S.; Anderson, J.; Goldberg, D.; Smolowitz, M.; C
 Proc. Natl. Acad. Sci. U.S.A. 86, 2463-2467, 1989
 A>Title: Cloning of a cDNA encoding the rat high molecular weight neurofilament peptide 22.
 A:Reference number: A32757; MUID:89184647
 A:Accession: A32757
 A>Status: Preliminary
 A:Molecule type: mRNA
 A:Residues: 559-566, 'E', 568-967, 'V', 969-987, 'GST', 1001-1022, 'E', 1024-1072 <LIE>
 A:Cross-references: GB:J04517; NID:9205679; PID:AAA41692.1; PID:9205680
 C:Superfamily: neurofilament triplet H protein
 C:Keywords: coiled coil; heterotrimer; intermediate filament; nerve; phosphoprotein

[illegible]

R:Breen, K.C.; Robinson, P.A.; Wion, D.; Anderson, B.H.

FEBS Lett. 241, 213-218, 1988

A:Title: Partial sequence of the rat heavy neurofilament polypeptide (NF-H). Identificat

A:Reference number: 502003; MIMD:89065087

A:Accession: 502003

A:Molecule type: mRNA

A:Residues: 1-854 <RE>

A:Cross-references: EMBL:X13804; NID:957828; PIDN:CAA32038.1; PID:957829

C:Superfamily: neurofilament triplet H protein

C:Keywords: coiled coil

Query Match 6.6%; Score 137.5; DB 2; Length 854;

Best Local Similarity 21.9%; Pred. No. 0.6;

Matches 94; Conservative 62; Mismatches 168; Indels 105; Gaps 20;

```

QY 4 PLRGPRSRGGGARRARAGRCRPAROSPARL-IPDTVLVDLVSDSEEV-----55
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 372 PVFAKSPAEAKSPASVSPG-EAKSPAEAKSPAEVKSPTATVKSFEAKSPAEVSPVTVK 430
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 56 --LEVADPVEVPVAPARLPAPAKPEQSDSDSEGAEGPAGAPRTLYRRRRRLDPGEAPV 113
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 431 SPDAKSPAEV-----KSPASVSPSEAKSPAGAKSPAEKSPVAKSPAEKSPAGAK- 484
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 114 VPVYSGVQSSSLNLPNNSLLKLPSEPEDEADLTN-----SGSSPSEDDA---LP 162
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 485 -PFAEAKSPAEAKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEK 538
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 163 SGSPWRKKLKKCKEKKKEEPPDDISPL-PPSSRNKSR-----KHTALQKLRVNVK 217
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 539 AKSPKAKKSPVKEIKPFAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEK 598
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 218 RLQDLRSCLSPKQHOSPALQSTDEEVVLVEGPVLPOSSRLFTLIRADLVRLP---VR 274
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 599 RPPDIR---SPEQVKSFAKEAKS-----PEKEETRTTEKVAKKKEVSPVEVK 645
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 275 MSEPLQNVVDMANHLGVSPNRIILFGESELSPATSTLKLGVADIIDCVLASSSEA 334
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 646 AKEP-----PKTV-----EKEKP-ATPKTE-----VKSRRDEA 674
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 335 TETISQELRLVQCKEKKOMLEISLSPDPLKVLMSHYEAMGSLGHLSPFDGTLKSLG 394
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 675 PKKQO-----KPKAKEKEPLTE--KPKDSPGEAKKEAKEKKAAP-----E 715
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 395 ELPADLGL 403
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 716 ETFAKLGAK 724

```

RESULT 5

MBEETE

C:Species: human nonstructural protein - human cytomegalovirus (strain Towne)

A:Note: host Homo sapiens (man)

C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 16-Jul-1999

C:Accession: A41808

R:He, Y.S.; Xu, L.; Huang, E.S.

J. Virol. 66, 1098-1108, 1992

A:Title: Characterization of human cytomegalovirus UL84 early gene and identification of

A:Reference number: A41808; MIMD:92114132

A:Accession: A41808

A:Molecule type: mRNA

A:Residues: 1-587 <HEV>

A:Cross-references: GB:M81432; NID:9330541; PIDN:AAA5947.1; PID:9330542

C:Superfamily: human cytomegalovirus 65k early nonstructural protein

C:Keywords: leucine zipper; nonstructural protein

F:114-135/Region: leucine zipper motif

F:325-373/Region: leucine zipper motif

Query Match 6.4%; Score 134.5; DB 1; Length 587;

Best Local Similarity 24.4%; Pred. No. 0.58;

Matches 105; Conservative 58; Mismatches 159; Indels 109; Gaps 22;

```

QY 6 RGRGPRSR-----GGGARARAGRCRPAROS-----PARLIPDTVLVDL--VSDSEEV 55
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 11 RARPRARRGGGGVGVSSRHS-GKCRORRALAPLFLATTTTMMGVASTDDDS 69
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 56 LEVADPVEVPVAPARLPAPAKPEQSDSDSEGAEGPAGAPRTLYRRRRRLDPGEAPV 115
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 70 LLKTPDEL-----DKHSSPQIL-----TLTKDINDIQR 101
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 116 VYSGK---VQSSNLIDN-----SLKLKCPSEPEDEADLTNGSSPSEDDALPS 163
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 102 VHRGTYHLIQHLDRPEELRDPQIILSTPLQGEANGESQ---TAPATSOEETAA-- 156
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 164 GSPRRKLLRKCKEKKKEEPPDDISPLPPOSSRNKSRHTALQKLRVNVKRLDLR 223
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 157 -----SHELEKKKEKEKEE--DED-----DRNDRER---GILCVSNEEDSVR 198
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 224 SCLS--FKQHOSPALQSTDEEVVLVEGPVLPOSSRLFTLIRADLVRLPVRSEPLON 281
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 199 PAFSLFPARPGCHILRSVIDQ-QLTMAIVRLSLNLEALRI-ITPPLKRVPLRRKAAHNT 256
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 282 VV-DHMANHLGVSPNRIILFGESELSPAT-PSLKL-GVADITDCVYLASSSEATETS 338
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 257 ALHDCMALHL-----PELTFESTLIDINNVTENAAAGVADAESTDADLT 299
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 339 QELRLRQCKEKKOMLEISLSPDPLKV-----LMSHYEAMGSLGHLSPFDGTLKSG 393
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 300 PTLTVRVRHAYCWHVRBGISGPRGLTSRISARLSETTKTLGPS-----VFGRLDLP 353
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 394 KELPADLGLS 404
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 354 NESPPDLTLLS 364

```

RESULT 6

T19431

hypothetical protein C25A1.10 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000

C:Accession: T19431

R:Mortimore, B.

submitted to the EMBL data library, October 1996

A:Reference number: T19431

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-971 <MIL>

A:Cross-references: EMBL:Z81038; PIDN:CAW0275.1; GSPDB:GN00019; CESP:C25A1.10

A:Experimental source: clone C25A1

C:Genetics:

A:Gene: CESP:C25A1.10

A:Map position: 1

A:Insertions: 38/3; 92/3; 201/3; 919/3

C:Superfamily: nucleolus-cytoplasm shuttle phosphoprotein

Query Match 6.3%; Score 133; DB 2; Length 971;

Best Local Similarity 23.3%; Pred. No. 1.3;

Matches 60; Conservative 35; Mismatches 115; Indels 48; Gaps 8;

```

QY 29 PRAROSPARLIPDTVLVDLVS---DSDEVLEVADPVEVPVAPARLPAPAKPEQSDSDSEG 85
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 231 PAAKPTPAKTPKRVVAKKAESSSDSDEKPPAKP--APAKATPKPAKKAADSSSDSD 288
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 86 ----AAGPAGAPPTLYRRRRRLD-----PGEAPVVPVYSGKVQSSSLNLI 128
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 289 DEAPAKKTPAKAAKPPVAKKAESSSDSDDEKPPAAKPTPAKATPKPAKKAESSSDSD 348
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 129 PDNSSLKLCIPSE--PPDEADLTNGSSPSEDNALPSGSPW--RKLKCKEKKKEEKEEF 185
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 349 DEKRPVAKPAPAKATPPVAKKAESSSDSDDEKPPAKPAPAKATPPVAKKAESSSDS 408
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 186 PDQDISPLPPOSSRNKS-----RKHTALQKLRVNVKRLQDLRSC 225

```

[illegible]

RESULT 7

Hypothetical protein C55C3.3 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: J29187
 R:Moessle, J., Stellyes, L.
 submitted to the EMBL Data Library, April 1996
 A:Description: The sequence of C. elegans cosmid C55C3.
 A:Reference number: Z20585
 A:Accession: J29187
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-792 <MOE>
 A:Cross-references: EMBL:U53335; PTDN:AAA96170.1; GSPDB:GN00022; CESP:C55C3.3
 A:Experimental source: strain Bristol N2; clone C55C3
 C:Genetics:
 A:Gene: CESP:C55C3.3
 A:Map position: 4
 A:Introns: 17/1; 46/3; 139/2; 173/1; 204/3; 250/3; 266/3; 286/2; 327/1; 362/3; 393/3; 43

Query Match	6.3%	Score	132.5	DB	2	Length	792
Best Local Similarity	20.98%	Pred. No.	1.1				
Matches	90	Conservative	50	Mismatches	135	Indels	155
						Gaps	19

```

0Y 11 RSRGRGRRRRRGARGRGRCPPARQSPARLIPDTVLVDVSDSEVLEADPVEVPARLP 70
Db 309 RRRGNNQIKKKKKPK--PLKRTP-----NNSDEKIKTKMSP----- 343
0Y 71 APAKEQSDSD-----SEGAEGPAGAPPTIVRRRRR-----LTDPG 109
Db 344 -PKRLQATDSDILQPLPSNNAMTVPEKSEANSPSSKINLVNRNGRIVNTDLKDCS 402
0Y 110 EAPVVPYISGKVOSSLNIPNSSLKLCISEPEDEADLTNCGSSPSEDALDPGSPWRK 169
Db 403 EARAP-----KC-DKLTDFOTSTWSESPA--DAISAVAP-TK 437
0Y 170 KLRRKCEKKEKKMEEPD--ODI-----SPLPPSSNKK----- 201
Db 438 LTGSOSOSEROETIEKMPNSFRELODSKFKPAEATIPKPEERKKGSTOSTTEPKLAGSQ 437
0Y 202 SRKHTLEOKLREVNKRLOD--LRSCLSPKHOSPALOSTDD----- 241
Db 498 SPSEKOEIOQMNSFRELODSKFKPAPSPKSEKPAQCSLSESPVPGNNRRNSETSQVD 557
0Y 242 -----EVLVEGVLPLOSSRLTLTKRCADLVRLPVMSSELPQNVYDH-----MAN 288
Db 558 TISPVPTLVGTGSOSEKOEIOQMNSFRELODSKFKPAQAPNKKPRVERKSGAEOGMS 617
0Y 289 HLGVPNRIILLFGSESLSPATPSTLK-----LGVADIIDCVYLA-----SS 331
Db 618 PLSTIS-----RAVAG-SPIAPKPRSPQAPLLETLANPPTIDAPTALITAIERSAEFSS 672
0Y 332 SEATTSOEL 341
Db 673 SHSMDPSNSL 682

```

RESULT

Protein: 65K early nonstructural protein - human cytomegalovirus (strain AD169)
N:Alternate names: UL84 protein
C:Species: human cytomegalovirus, human herpesvirus 5
A:Note: host Homo sapiens (man)

C>Date: 31-Dec1992 #sequence_revision 31-Dec-1992 #text_change 16-Jul-1999
C:Accession: S09848
R:Chae, M.S.; Bankier, A.T.; Beck, S.; Bohm, R.; Brown, C.M.; Cerny, R.; Horsnell, T.
M.; Barrell, B.G.
Curr. Top. Microbiol. Immunol. 154, 125-169, 1990
A>Title: Analysis of the protein-coding content of the sequence of human cytomegalovirus
A:Reference number: S09749; MUID:90265039
A:Accession: S09848
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-586 <CH>
A:Cross-references: EMBL:X17403; NID:G59591; PID:CA45358.1; PID:Q1780862
A>Note: this sequence was submitted to the EMBL Data Library, December 1989
C:Keywords: leucine zipper; nonstructural protein
F:114-135/Region: leucine zipper motif
F:324-372/Region: leucine zipper motif

Query Match	6 3%	Score 132;	DB 1;	Length 586;
Best Local Similarity	23.6%	Pred. NO. 0.82;		
Matches 104;	Conservative 62;	Mismatches 145;	Indels 130;	Gaps 25;

```

Query Match Score 6.3%; DB 1; Length 586;
Best Local Similarity 23.6%; Pred. No. 0.82;
Matches 104; Conservative 62; Mismatches 145; Indels 130; Gaps 25

OY      6 RGRGPRSR-----GGRGARARAGRGRCPPAROS----PARLPDYLVVL--YSDSDELY 55
          | | | | |         : : : : :       | : | : | : | : | : | : | : | : |
Db      11 RRNRPPARRGGGGVGSNSRRS-GCGRKORRLSLAPLFLTLATTTTTTMMGAASDDDS    69
          | | | | |         : : : : :       | : | : | : | : | : | : | : | : |
OY      56 LEVADVEVPVARLPAPAKPEQDSDSDSGAAGEPAGADRTLVRRRRRLLDGEAPVP   115
          | | | | |         : : | : | : | : | : | : | : | : | : | : | : |
Db      70 LILKTPDEL-----DKYSSPOLII-----TLTKHDIHQRP    101
          | | | | |         : : | : | : | : | : | : | : | : | : | : | : |

```

```

QY 116 VYSGK---VOSSLNIPDN-----SLLIKCPSPEDEADLTNGSSPSSEDLAPS 163
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 102 VARGTYHLIQLHLDRLPEELRDPQIILSTPELOL--GEANDESQ--TAPATLDEEETAAS 157
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 164 GSPMFKLLRKCKEKEKKEEFPRODIDPRLQPSRRNKSKNHTALQALREYKRRLODLR 223
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 158 HEPERKK-----EKEKKEE---DED-----DRDNRER-----GILCVSNEDSDVR 197
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 224 SCLTS--PKOHOSPALOSIDDEVILVEGCVLPSSRFLTKICRADAIVLRVPRNSEPTLN 281
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 198 PAFSLFPRAPRGCHILRSYIDO-QLTRMAIVRLSLNLFALRI-TPLPLKRLPRLR----RK 250C
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 283 VVDHANHLGVSFNRILLFGESELSPATPSTYKLGVAIDIDCVVLASSGEATQSQEL 341
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 251 AANHNLALH-----DCLALHLEPFLPEPLLDLNNNTENNASYAD---TAESTDA-DLPTL 300
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 342 RLVQGEKKHOMLETISLSPDPLKVLMSHYEAMGLSG-----HKLS----- 383
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 302 TVRVRR-----HALCHNRVEG--GISGPRCLTISRISARLSETTAKTLGP 342
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 384 FFFDGTKLSGKELPADLGLES 404
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 343 SVFGRIELDPNESPDPDLTLLS 363
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 9

199696
 nucleolin homolog num1 - alfalfa
 N:Alternate names: num1 protein
 C:Species: Medicago sativa (alfalfa)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
 C:Accession: T09648
 R:Boegre, L.; Jonak, C.; Mink, M.; Mesiklene, I.; Traas, J.; Ha, D.T.C.; Swoboda, I.;
 Plant Cell 8, 417-428, 1996
 A:Title: Developmental and cell cycle regulation of Alfalfa nucleos1 a plant homolog of
 A:Reference number: Z16796; MUID:96361876
 A:Accession: T09648
 A:Status: preliminary; translated from GR/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-635 <BOE>
 A:Cross-references: EMBL:X88845; NID:q1279562; PIDN:CAA61298.1; PID:q1279563
 C:Genetics:


```
Db 173 AAEKL-----NIPDYQEKEDAFMQLVPPAV-----GAD 200
Qy 268 LVRLPYRMSEPLQNVVDMANHLGVSPNR---ILLFGESELSPTATPSTLKLGVADIID 324
Db 201 IMNLFPEKSAALM-----YHIGAMPKARQLLAMDGQSAL----- 235
Qy 325 CVALSSSEATEFTSÖELRLRVGCKEKHÖMLEISLSP--DSPL 364
Db 236 -----IELTRLSERLTLKPRGCK-----QLSSAPPADQPI 264

RESULT 12
S14871
suppressor two of zeste protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 21-Jul-2000
C:Accession: S14871; S16845
R:Brunk, B.P.; Adler, P.N.
submitted to the EMBL Data Library, November 1990
A:Description: The Drosophila regulatory gene suppressor two of zeste encodes a large z
A:Reference number: S14871
A:Accession: S14871
A:Molecule type: mRNA
A:Residues: 1-1365 <BRU>
A:Cross-references: EMBL:X56799; NID:g8528; PID:g8529
R:Brunk, B.P.; Adler, P.N.
Nucleic Acids Res. 19, 3149, 1991
A:Title: The sequence of the Drosophila regulatory gene suppressor two of zeste.
A:Reference number: S16845; MUID:91279476
A:Accession: S16845
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-612,614-784,'N',786-830,'R',832-965,967-1064,'E',1066-1096,'N',1097-1286,'
A:Cross-references: EMBL:X56798; NID:g8526; PIDN:CAA0134.1; PID:g8527
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1990
C:Genetics:
A:Gene: FlyBase:Su(2)2
A:Cross-references: FlyBase:FBgn0008654
A:Introns: 55/1; 139/2; 219/3
C:Superfamily: Drosophila suppressor protein of zeste; RING finger homology
C:Keywords: DNA binding; nucleus; zinc finger
F:31-79/Domain: RING finger homology <RNG>

Query Match 6.1%; Score 127; DB 2; Length 1365;
Best Local Similarity 23.1%; Pred. No. 4.6;
Matches 84; Conservative 47; Mismatches 125; Indels 108; Gaps 16;

Qy 110 EAPVVPYYSKVOSSLNLDNSSLKLCPS-EPDEADITNGSSP--SED--DALPSG 164
Db 220 ESPMAFCYRILLYDNEQTKDNNLSRINDIEPHSVRSKSAKSVTFADLESEIDSG 279
Qy 165 SPNRKRLRKCKEKEKMEFPDODISLPPOSSRNK---SRKHTEA-----LQKRE 214
Db 280 SP-RSKYRCK-----TPPKVSPSKNKRLLTSKREAPESPVPNFKSLRS 323
Qy 215 VNRRLQDL-----RSCUSPKQHGSPALQSTDEVVLEGVLPQS---SRLETL 260
Db 324 NDMRYSDAVASKVSEPEQOFLPREREQOPLFANTNIIVSIPQSLRKSVDYADEDFEL 383
Qy 261 KIKCRADLVLP-----VRMSEPL-----QNVVDIMANHLGV 292
Db 384 KLANRKRGVGHLPKILKELNSMKSLMPLSAGPRLLEDTSOSSSCSAOQDLLEYAKNIGL 443
Qy 293 SPNRILLFG-----ESELPTATP-----STLKLGVADIDCVLASSEATETS 338
Db 444 KPIEOPLOQASNPDKSYSPNAPSMSCSSSTNGSSSLTAD-----ASTSTSSSS 496
Qy 339 QEDRLR-----VQCKEKHÖMLEISLSPSLKVLMSHYEAMGLSGHKLFFFDGTR 390
Db 497 HRRKKKHSKEPKDANGKRRKRLHAEISQTDGKKMKVIT-----AKPNKLDLPKRSLSL 550
Qy 391 LSGK 394
```

```
Db 551 ASGE 554
||:
||:
RESULT 13
T16543
hypothetical protein K03C7.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Mar-2000
C:Accession: T16543
R:Leinbach, D.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid K03C7.
A:Reference number: Z18532
A:Accession: T16543
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-856 <LEU>
A:Cross-references: EMBL:U40059; NID:g1055170; PID:g1055171; PIDN:AAA81138.1; CESP:K0
C:Genetics:
A:Gene: CESP:K03C7.1
A:Introns: 176/1; 632/2; 717/3; 746/3; 791/3
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo

Query Match 6.0%; Score 126; DB 2; Length 856;
Best Local Similarity 21.4%; Pred. No. 3;
Matches 85; Conservative 55; Mismatches 153; Indels 104; Gaps 17;

Qy 2 AEPLRGRGPRSGRGARARGRCPRAROSPRLIPDTLVLDVSPSDEVLVEVADP 61
Db 382 ADPL--GGPKKDPKLAKKAPAKK--PTDKRPK-----DAPKDAKPTPEEPAPK 427
Qy 62 V-----EVPARLPAPAKPEODSDSGAEGPAGA-----PRTLYRR 100
Db 428 VAPKKKPPWEEDPDPEPEADFTMPAKKPDTEPAD-----PLGCKNKDPKLAKKA 482
Qy 101 RRRRLID---PGAPVVPYYSKVOSSLNLDNSSLKLCPS-EPDEADITNGSSP--SED--DALPSG 164
Db 483 PTKKPADKKPKSEEPKPYAPKWKMPWEEDPDPEADFTYPIKGEDEDEPEADDEE 542
Qy 154 SPSEDALTSGSPWRKLLKCKEKKMEFPDODISLPQ----- 196
Db 543 EPDE---PAEDEPKKKKKRKHKKRPKPKVPEPEKEPTPEVPVPAKWTAPLKPEE 599
Qy 197 -----SSRNKSRKHTEALQKLRVNRKLDLRLSCLSPKHOSPAL----- 236
Db 600 PTPMPKKEKTIERNKEERILPALRYAKK-PRELVYIIFVLPWE-QTALITTOEGMAF 657
Qy 237 ---QSTDEVVLEGVLP--OSSRLFTL-----KIKCRADLVLPVMSSEPLQNVVDH 285
Db 658 GKSRANVENVNGDKPIYOGAVDSKTVIPLMNDSCANRSGMTAFGA-PREIDQNVVDH 716
Qy 286 MANHLGVSPNRILLFGESELSPTATPSTLKLGVADI 322
Db 717 ---HKGSOGIILPLAKGTTHPHGEGYTIIRQTADV 750

RESULT 14
JW0057
gravin - human
C:Species: Homo sapiens (man)
C:Date: 13-Jun-1998 #sequence_revision 10-Jul-1998 #text_change 08-Oct-1999
C:Accession: JW0057
R:Sato, N.; Kokame, K.; Shimokado, K.; Kato, H.; Miyata, T.
J. Biochem. 123, 1119-1126, 1998
A:Title: Changes of gene expression by lysophosphatidylcholine in vascular endothelia
A:Reference number: JW0057; MUID:98269042
A:Molecule type: mRNA
A:Residues: 1-1684 <SAT>
A:Cross-references: DBJ:AB003476; NID:g2081606; PIDN:BAI19927.1; PID:d1020716; PID:g
C:Comment: This protein regulates cell growth.
```


F:433-439/Region: nuclear location signal
 F:522-527/Region: nuclear location signal
 F:591-596/Region: nuclear location signal
 F:671-676/Region: nuclear location signal

Query Match 5.9%; Score 124.5; DB 2; Length 1684;
 Best Local Similarity 20.7%; Pred. No. 8.5;
 Matches 94; Conservative 70; Mismatches 171; Indels 119; Gaps 21;

QY 43 VLVDVSDSEVLEVADVEVPAVRLPAPAKPEQSDSD----- 82
 Db 32 VVHDTDDGGEETPEIIEQIPSSSENLLETQPTESQANDIGFKVEKVEGKFTVKDK 91
 QY 83 -----SEGAAGPAGA-----PRTLVRRRRLDPGEAPVVPYSGK-VQSS 124
 Db 92 TEKPDYQVLLTVKKDEGEAGAGDHDPSLAGEAASKESEPQSTETKPEETLKREQSH 151
 QY 125 LNLIP--DNSSLILKICSEPED--EADLTNSGSSPEDDALPSGSP-----WRK 169
 Db 152 AEISPPAESGQAVECKEKEGEKQEKESKSAESPTSPVISEGTPKFKFTQGMAGMRK 211
 QY 170 K--LRKCKEK-----EKKMEFPDODISPLPQSSRNKSRKHTALQKREVNKRLODLR 223
 Db 212 KTSFRKREDEVEASEKKKEQEPK-----VDTEEDGKAEVASEKLTASEQAHQ-EPAE 265
 QY 224 SCLSPK---OHOSPALOSTDDEVVLVEGP---VLPOSSRLFTLKINGRADVRLPYRMS 276
 Db 266 SAHPRRLSAEYKVELRS-EQVVSQSGPSEKPAPLATEYFDEKIEYHOEEVVAEVAHS 324
 QY 277 EPLQNVVDHMANHLGVSPNRIILLGSELSPTATPTLKLGVADIIDCVVLASSSEATE 336
 Db 325 TVEERTEQ-----KTEVEETA-----GSVPAEELVEMDAEPQAE 360
 QY 337 TSQEL-RLR---VOGKEKHMLEISLSPDPL-----KVLMSHYEAMGLSGHKL 382
 Db 361 PAKELVLIKETCVSGEPTQAD--LSPDEKVLSPPEGVSEVEMLSQGRMKVQGSPL 418
 QY 383 SFFPDGT---KLSGKELPADLG---LESGLIEV 410
 Db 419 KKLFTSTGLKLSGKKOKGKRGGDESGEHTQV 452

RESULT 15

T02345
 hypothetical protein KIAA0324 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 05-Nov-1999

C:Accession: T02345

R:Ricke, D.O.; Bruce, D.; Mundt, M.; Doggett, N.; Munk, C.; Saunders, E.; Robinson, D.;
 re, J.; White, S.; Ueng, S.; Tatum, O.; Campbell, C.; Fawcett, J.; Deaven, L.
 submitted to the EMBL Data Library, March 1998

A:Description: Sequencing of human chromosome 16p13.3.

A:Reference number: Z14664

A:Accession: T02345

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1791 <R10>

A:Cross-references: EMBL:AC004493; NID:g2996648; PIDN:AA08453.1; PID:g2996650

C:Genetics:

A:Map position: 16

A:Introns: 1610/2; 1706/2

A:Note: KIAA0324

Query Match 5.9%; Score 124.5; DB 2; Length 1791;
 Best Local Similarity 25.3%; Pred. No. 9.2;

Matches 72; Conservative 26; Mismatches 96; Indels 91; Gaps 13;

QY 34 SPARLPDFTVLVDVSDSEVLEVADVEVPAVRLPAPAKPEQSDSDSEGAAGPAGA 93
 Db 414 SSELSPDAVEKAGWSSNQ-----SISSPVDVAVPRTSPRER-----SSSASSPMKDGL 463

QY 94 PRTLVRRRRLDPCGAPVVPYSGKVOSSLNLPDNNSTL-----LKLCPSEPD 144
 Db 464 PRTSRRSR-----SGSSPGLRDGG-----TPSRHLSGSSPGMKDIPRTPSRGRS 510
 QY 145 EADLTNSGSSPEDDALP-----SGSPWRKRLKCK--EKEKKMEFPDQ----- 188
 Db 511 ECD-----SSP-EPKALPQPRPRSRSPSPPELNKCLTPQRRSGSSSVQKTVARTP 564
 QY 189 -----DISLPQPSR-----MKSRKHTALQKREVNKRLODLRSCLS 227
 Db 565 LQGRSRGSSQELDVKPSASPOERSESDSDPSAKTRTPLQRSRSGSSPEVDSKSRLS 624
 QY 228 PKQHOS-----PALOSTDDEVVLVEGP---VLPOSSR 256
 Db 625 PRSRSGSSPEVKDKPRAAPRAQSGSDSSPEPKAPAPRALPRSR 669

Search completed: April 22, 2002, 10:07:02
 Job time: 47 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 22, 2002, 10:06:15 ; Search time 13.21 Seconds
(without alignments)
1143.521 Million cell updates/sec

Title: US-09-617-923-2
Perfect score: 2099
Sequence: 1 MAEPLRGGRSGRGARR.....GKELPADLGESGLIEWWG 412

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 3664827 residues
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	135.5	6.5	831	1 NFN_RAT	P16884 rattus norv
2	134.5	6.4	587	1 ULR4_HCMVT	P29839 human cytom
3	132	6.3	586	1 ULR4_HCMVA	P16727 human cytom
4	128.5	6.1	704	1 NP14_RAT	P41777 rattus norv
5	128	6.1	303	1 VG08_BP22	P26748 bacterioph
6	127	6.1	1365	1 SUZ2_DROME	P25172 drosophila
7	122.5	5.8	1781	1 AKAC_HUMAN	Q02952 homo sapien
8	121	5.8	1109	1 TCF8_RAT	Q62947 rattus norv
9	119	5.7	1603	1 PSC_DROME	P35820 drosophila
10	116.5	5.6	1039	1 MS11_DROME	P50535 drosophila
11	116.5	5.6	1061	1 LBS_DROME	P34739 drosophila
12	116	5.5	1043	1 TCF8_MESAU	Q60542 mesocricetu
13	115.5	5.5	872	1 SCD5_YEAST	P34758 saccharomyc
14	114.5	5.5	1411	1 TCOF_HUMAN	Q14428 homo sapien
15	113.5	5.4	1280	1 DYNA_RAT	P28023 rattus norv
16	112.5	5.4	95	1 SM32_HUMAN	P55855 homo sapien
17	112	5.3	919	1 DNLI_HUMAN	P18858 homo sapien
18	112	5.3	943	1 CENC_HUMAN	Q03188 homo sapien
19	112	5.3	1004	1 PRPX_RAT	Q64475 rattus norv
20	111.5	5.3	1170	1 XPE_MOUSE	P19369 mus musculu
21	111	5.3	741	1 BSG2_DROME	P11929 drosophila
22	111	5.3	897	1 EP15_MOUSE	P42567 mus musculu
23	111	5.3	1281	1 DYNA_MOUSE	Q08788 mus musculu
24	110.5	5.3	919	1 SYPN_HUMAN	Q9465 homo sapien
25	110.5	5.3	2517	1 NCR2_HUMAN	Q94618 h nucleat r
26	110	5.2	542	1 TUL1_HUMAN	Q00294 homo sapien
27	110	5.2	612	1 ARRS_MAIZE	P13027 zea mays (m
28	110	5.2	883	1 E74B_DROME	P11536 drosophila
29	110	5.2	1206	1 FM14_MOUSE	Q05859 mus musculu
30	110	5.2	2150	1 SDC3_CAEBL	P34706 caenorhabdi
31	109.5	5.2	536	1 GAG_MVCB	P27460 cas-br-e mu
32	109.5	5.2	972	1 ORF4_SCHRO	Q97794 schizosach
33	109	5.2	974	1 YMB4_CAEBL	Q03601 caenorhabdi

34	109	5.2	1089	1 Y553_HUMAN	O9ukf3 homo sapien
35	108.5	5.2	487	1 ATR2_CHICK	O93602 gallus gall
36	108.5	5.2	505	1 TUB_MOUSE	P50586 mus musculu
37	108.5	5.2	1117	1 TCF8_MOUSE	Q64318 mus musculu
38	108	5.1	487	1 ATR2_HUMAN	P15336 homo sapien
39	108	5.1	487	1 ATR2_MOUSE	P16951 mus musculu
40	108	5.1	673	1 FXR2_HUMAN	P51116 homo sapien
41	108	5.1	1402	1 IF4G_RABIT	P41110 oryctolagus
42	107.5	5.1	532	1 ICP0_HSVB	P28890 equine herp
43	107	5.1	487	1 ATR2_RAT	Q00969 rattus norv
44	107	5.1	650	1 NDCU_XENLA	P20397 xenopus lae
45	107	5.1	1224	1 DYNA_CHICK	P35458 gallus gall

ALIGNMENTS

RESULT	ID	NEF_RAT	STANDARD	PRT	831 AA.
AC	P16884	063368;			
DT	01-AUG-1990	(Rel. 15, Created)			
DT	01-FEB-1994	(Rel. 28, Last sequence update)			
DT	20-AUG-2001	(Rel. 40, Last annotation update)			
DE	NEUROFILAMENT TRIPLET H PROTEIN (200 KDA NEUROFILAMENT PROTEIN)				
DE	(NEUROFILAMENT HEAVY POLYPEPTIDE) (NF-H) (FRAGMENT).				
GN	NEFH OR NEFH.				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.				
OX	NCBI_TaxID=10116;				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Brain;				
RX	MEDLINE=89065087; PubMed=3143606;				
RA	Breen K.C., Robinson P.A., Wion D., Anderson B.H.;				
RT	"Partial sequence of the rat heavy neurofilament polypeptide (NF-H).				
RT	Identification of putative phosphorylation sites."				
RL	FEBS Lett. 241:213-218(1988).				
RN					
RP	SEQUENCE OF 37-831 FROM N.A.				
RX	MEDLINE=88309090; PubMed=2457365;				
RA	Dautigny A., Pham-Dinh D., Rousset C., Felix J.M., Nussbaum J.L.,				
RT	Jolles P.;				
RT	"The large neurofilament subunit (NF-H) of the rat: cDNA cloning and				
RT	in situ detection."				
RL	Biochem. Biophys. Res. Commun. 154:1099-1106(1988).				
RN					
RP	SEQUENCE OF 1-89 AND 243-313 FROM N.A.				
RX	MEDLINE=87080760; PubMed=2878828;				
RA	Robinson P.A., Wion D., Anderson B.H.;				
RT	"Isolation of a cDNA for the rat heavy neurofilament polypeptide				
RT	(NF-H)."				
RT	FEBS Lett. 209:203-205(1986).				
RL					
RN					
RP	SEQUENCE OF 318-831 FROM N.A.				
RX	MEDLINE=89184647; PubMed=2928342;				
RA	Lieberburg I., Spinner N., Snyder S., Goldhaber D.,				
RA	Smolowitz M., Carroll Z., Emanuel B.S., Bretiner J., Rubin L.;				
RT	"Cloning of a cDNA encoding the rat high molecular weight				
RT	neurofilament peptide (NF-H): developmental and tissue expression in				
RT	the rat, and mapping of its human homologue to chromosomes 1 and				
RT	22."				
RL	Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).				
CC	-1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,				
CC	AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.				
CC	NEF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT				
CC	SUBSERVED BY THE TWO SMALLER NF PROTEINS.				
CC	-1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPLET K-S-P, NEF IS				
CC	PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS				
CC	THOUGHT THAT PHOSPHORYLATION OF NEF RESULTS IN THE FORMATION OF				
CC	INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE				
CC	OF AXONAL CALIBER.				

FT REPEAT 524 535 ACIDIC SERINE CLUSTER 10.
 FT REPEAT 559 570 ACIDIC SERINE CLUSTER 11.
 FT MOD_RES 567 567 PHOSPHORYLATION (BY CK2).
 FT VARIANT 150 150 MISSING (IN NOP140B).
 SQ SEQUENCE 704 AA; 73562 MW; 14DF1BF2DE483EA3 CRC64;

Query Match 6.1%; Score 128.5; DB 1; Length 704;
 Best Local Similarity 23.2%; Pred. No. 0.8;
 Matches 71; Conservative 35; Mismatches 105; Indels 95; Gaps 12;

OY 33 GSPALIPDTVYLVDSDE-----VLEVADPYEVVPAARL---PAPAKPQSDSDSSE 84
 DB 317 GSPKKAAGTADPADSSADSSSESSSESEKTPAKTYVSKTPAPAPYKKAESSDSS 376
 OY 85 GAEGPAGAPRTLVRRRRRLDPGEAPVYV-----YSGKVOSSLNL 127
 DB 377 DSDSDEDEAPAPVATKSPKPAVTPKPAKAVATPKQAGSGQKQKADSSSE 436
 OY 128 IPDNSS-----LKLCPSEPDDEA-----DLTNGSSPSSEDD----- 159
 DB 437 EESSSESEETATKSVTPKARVYAKAAPSLPAKQAPRAGGSDSSSESSSESEKTPPK 496
 OY 160 -----ALPSGSPMKKRLK-----KCEKEKKMEFPDDISPLQ----- 195
 DB 497 PPAKKAAGAAVPPKTPVKKAAESSSSSSSESDSSSESEKTPSKATPKQAGKA 552
 OY 196 ---PSSRN-KSRKHTALQKREVNKRLQDLRSCLSPKQHOSPALQSTDEVVLVEGPVL 251
 DB 553 NGVPSAQNKAKKESEEEDEDTQKKAAGTKPGSKKKKKHH-----ETPDEA-----AT 602
 OY 252 PQSSRL 257
 DB 603 PQSKRV 608

RESULT 5
 VC08_BPP22 STANDARD; PRT; 303 AA.
 AC P26748;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE SCAFFOLDING PROTEIN (PROTEIN GP8).
 GN 8.
 OS Bacteriophage P22.
 OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Podoviridae.
 OX NCBI_TaxID=10754;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-15.
 RX MEDLINE=91306435; PubMed=1853558;
 RA Epipler K., Wyckoff E., Goates J., Parr R., Casjens S.;
 RT "Nucleotide sequence of the bacteriophage P22 genes required for DNA packaging.";
 RT Virology 183:519-538(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Kropinski A.M.B., VanderByl C.S.;
 RT "The completed sequence of genome of Salmonella phage P22.";
 RT Submitted (DEC-1999) to the EMBL/genbank/DBJ databases.
 CC -1- FUNCTION: REQUIRED FOR SUCCESSFUL CONDENSATION OF DNA WITHIN THE
 CC CAPSID. THE INTERIOR OF THE PROHEAD IS FILLED WITH THE GP8
 CC PROTEIN. THE SCAFFOLDING PROTEIN IS LOST FROM THE STRUCTURE
 CC DURING PACKAGING.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
 CC -1- INDUCTION: THE SCAFFOLDING PROTEIN NEGATIVELY REGULATES ITS OWN
 CC SYNTHESIS WHEN IT IS NOT ASSEMBLED INTO PROHEADS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.
 CC -----
 CC EMBL: M59749; AAA72962.1;
 CC DR EMBL: AF217253; AAF75046.1;
 CC DR PIR: D40474; Z8BP22.
 CC KW late protein.
 CC SEQUENCE 303 AA; 33564 MW; B86BD625D5011ABE CRC64;

Query Match 6.1%; Score 128; DB 1; Length 303;
 Best Local Similarity 21.3%; Pred. No. 0.31;
 Matches 73; Conservative 49; Mismatches 100; Indels 120; Gaps 15;

OY 41 DTIVLDVSDS--DEVLVADPYEVVPAARLPAAPAKPQSDSDSSEGAEGPAGAPRTLV 98
 DB 25 DSLVVDNANDNAGGEGFEIY-----LKDDETAAPKQDPAKKAEEFARRRIE 69
 OY 99 RRRRRRLDPGAPAVPYVYSGKVOSSLNLIPDNSSLKLCSEPDDEADLTNGSSPSSE 158
 DB 70 RRQRELEQOMEA---VKRGELPSLVNPD-----LPPQDINVLSEGLAKYDY 118
 OY 159 D---AL---PSSGSPMKRLK---KCEKEKKMEFPDDISPLPQSSRNKSRKHTTE 207
 DB 119 DNSRLALAFNMANTMLKADARSNAAVEGKTKQETQOQAQV-----EAAKHYD 172
 OY 208 ALQKREVNKRLQDLRSCLSPKQHOSPALQSTDEVVLVEGPVLPQSSRLFTLKIRCAD 267
 DB 173 AAELK-----NIPDYQEKEDAFMQLVPPAV-----GAD 200
 OY 268 LVRLVRRSEPLQNVVDHMANHLGVSPNR---ILLFGESELSPTATPSTLKGADVIID 324
 DB 201 IMRLPEKSAALM-----YHLGANPEKARQLAMDQGSAL----- 235
 OY 325 CVVLASSEATETSOELRLRVQKEKHOMLEISLSP--DSP 364
 DB 236 -----IELRLSERLTLKPRGK-----QISSAPPADQPI 264

RESULT 6
 SUZ2_DROME STANDARD; PRT; 1365 AA.
 ID SUZ2_DROME
 AC P25172;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE SUPPRESSOR 2 OF ZESTY PROTEIN (PROTEIN POSTERIOR SEX COMBS).
 GN SU(2).
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CANTON-S.
 RX MEDLINE=91279476; PubMed=2057369;
 RA Brunk B.P., Adler P.N.;
 RT "The sequence of the Drosophila regulatory gene Suppressor two of
 RT zeste.";
 RT Nucleic Acids Res. 19:3149-3149(1991).
 CC -1- FUNCTION: REGULATES EXPRESSION OF THE HOMEOTIC SELECTOR GENES BY
 CC INFLUENCING HIGHER-ORDER CHROMATIN STRUCTURE THROUGH INTERACTION
 CC WITH OTHER PROTEINS.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
 CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch).

```

CC -----
DR EMBL: X56798; CAA40134.1; -
DR EMBL: X56799; CAA40135.1; -
DR PIR: S16845; S16845.
DR PIR: S14871; S14871.
DR Flybase: FBgn0008654; Su(z)2.
DR InterPro: IPR001841; Znf_ring.
DR Pfam: PF00097; zfc3hc4.1.
DR SMART: SM00184; RING.1.
DR PROSITE: PS00518; ZINC_FINGER_C3HC4.1.
KM Zinc-finger: Developmental protein; DNA-binding; Nuclear protein.
FT ZN_FING 35 73
FT DOMAIN 623 628 POLY-GLN.
FT DOMAIN 1077 1096 POLY-ASN.
FT DOMAIN 1241 1251 POLY-SER.
FT CONFLICT 603 603 MISSING (IN REF. 1; CAA40134).
FT CONFLICT 785 785 K -> N (IN REF. 1; CAA40134).
FT CONFLICT 831 831 A -> R (IN REF. 1; CAA40134).
FT CONFLICT 965 965 MISSING (IN REF. 1; CAA40134).
FT CONFLICT 1065 1065 D -> E (IN REF. 1; CAA40134).
FT CONFLICT 1076 1076 MISSING (IN REF. 1; CAA40134).
FT CONFLICT 1287 1287 A -> P (IN REF. 1; CAA40134).
SQ SEQUENCE 1365 AA; 146058 MW; 7B4B0F35B0FA683 CRC64;

```

Query Match 6.1%; Score 127; DB 1; Length 1365;

Best Local Similarity 23.1%; Pred. No. 2.2; Mismatches 125; Indels 108; Gaps 16;

```

Matches 84; Conservative 47; Mismatches 125; Indels 108; Gaps 16;

QY 110 EAPVVPVYSGKVOSSLNLIPDNSSLLKICPS-EPEDADLTJNSGSSP--SED--DALPSG 164
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 220 ESMACVYRLLLYDNDQTKNDENRLSRINODEPEHVSRRSKAKSVTFADLESEIDSG 279
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

QY 165 SPWRKRLRKCKEKEKMEFPDQDISPLPQSSRNK-----SKKHEDA-----LQKRE 214
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 280 SP-RSVVRCK-----TPPVSPSSKKNKRLTSSKREAPESPVSNFSLRS 323
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

QY 215 VNRKLQDL-----RSCLSKPOHSPALQSTDEVLVEGVPYLPOS---SRLETL 260
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 324 NDMRYSDVAVSKYKSPEDQEQFLPREQOQLPEANTNIVYISIPQQLKRSYDADDEL 383
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

QY 261 KTRCADLVRP-----VRMSEPL-----QNVVDHMANHLGV 292
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 384 KLANRKGVGHFLPKLIELNSMKSLSMPLSAGRLDQTSKSSCSAQDLDEYAAKNI 443
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

QY 293 SPRIILLFG-----ESELSPATNP-----STLKGVAIIICVYLASSEATERS 338
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 444 KPIEQLQOSASNPDSKYSPPNAPSMSSSSTNGSSSSSLGTAD-----ASTSTSTSS 496
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

QY 339 QELRLR-----VQKEKHQMLEISLSPDPLKYLAMSHYEERAMGLSGHKLSFFPDG 390
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 497 HRRKKKKHSEKPRDANGKRRKKLHAELSSQTDGKMKVKIT-----AKPNKLDLFRKSHSL 550
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

QY 391 LSGK 394
   1 1
DB 551 ASGE 554

```

RESULT 7

```

AKAC_HUMAN STANDARD; PRT; 1781 AA.
AC Q02952; Q09970; Q00498; Q00310;
DT 01-FEB-1994 (Rel. 28; Created)
DT 20-AUG-2001 (Rel. 40; Last sequence update)
DT 20-AUG-2001 (Rel. 40; Last annotation update)
DE A-KINASE ANCHOR PROTEIN 12 (A-KINASE ANCHOR PROTEIN 250 KDA) (AKAP
GN AKAP12 OR AKAP250.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
NCBI_TaxID=9606;
RN [1]

```

```

RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Heart;
RX MEDLINE=97153077; PubMed=9000000;
RA Nauert J.B., Klauck T.M., Langeberg L.K., Scott J.D.;
RT "Gravin, an autoantigen recognized by serum from myasthenia gravis
RL patients, is a kinase scaffold protein.";
RL Curr. Biol. 7:52-62(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=umbilical vein endothelial cells;
RX MEDLINE=98269042; PubMed=9604001;
RA Sato N., Kokame K., Shimokado K., Kato H., Miyata T.;
RT "Changes of gene expression by lysophosphatidylcholine in vascular
RT endothelial cells: 12 up-regulated distinct genes including 5 cell
RT growth-related, 3 thrombosis-related, and 4 others.";
RL J. Biochem. 123:1119-1126(1998).
RN [3]
RP SEQUENCE OF 43-1781 FROM N.A.
RC TISSUE=umbilical vein endothelial cells;
RX Bowditch R.D., Ginsberg M.H.;
RA "Sequence of gravin cDNA isolated from a human umbilical vein
RT endothelial cell library.";
RL Submitted (Apr-1997) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1477-1781 FROM N.A.
RC TISSUE=umbilical vein endothelial cells;
RX MEDLINE=92395179; PubMed=1522245;
RA Gordon T., Grove B., Loftus J.C., O'Toole T., McMillan R.,
RA Lindstrom J., Ginsberg M.H.;
RT "Molecular cloning and preliminary characterization of a novel
RT cytoplasmic antigen recognized by myasthenia gravis sera.";
RL J. Clin. Invest. 90:992-999(1992).
CC -1- FUNCTION: ANCHORING PROTEIN THAT MEDIATES THE SUBCELLULAR
CC COMPARTMENTATION OF PROTEIN KINASE (PKA) AND PROTEIN KINASE C
CC (PKC).
CC -1- SUBUNIT: BINDS TO DIMERIC RII-ALPHA REGULATORY SUBUNIT OF PKA.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. MAY BE PART OF THE CORTICAL
CC CYTOSKELETON.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: 1 (SHOWN HERE) AND 2; MAY BE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ENDOTHELIAL CELLS, CULTURED
CC FIBROBLASTS AND OSTEOSARCOMA, BUT NOT IN PLATELETS, LEUKOCYTES,
CC MONOCYTIC CELL LINES OR PERIPHERAL BLOOD CELLS.
CC -1- INDICATION: ACTIVATED BY LYSOPHOSPHATIDYLCHOLINE (LYSOPC).
CC -1- DOMAIN: POLYBASIC REGIONS LOCATED BETWEEN RESIDUES 265 AND 556 ARE
CC INVOLVED IN BINDING PKC.
CC -1- DISEASE: ANTIBODIES TO THE C-TERMINAL OF GRAVIN CAN BE PRODUCED BY
CC PATIENTS WITH MYASTHENIA GRAVIS (MG).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U81607; AAC51366.1; -
DR EMBL: AF001504; AAB58838.1; -
DR EMBL: AB003476; BAA19927.1; -
DR EMBL: M96322; AAA35931.1; -
DR PIR: A43922; A43922.
DR MIM: 604698; -
DR InterPro: IPR001573; PkinA_anch.
KM Antigen; Alternative splicing.
FT DOMAIN 603 633 AKAP 1.
FT DOMAIN 752 782 AKAP 2.
FT DOMAIN 797 827 AKAP 3.
FT DOMAIN 98 101 POLY-GLU.
FT DOMAIN 265 556 INVOLVED IN PKC-BINDING (PROBABLE).
FT DOMAIN 1540 1553 RII-BINDING (PROBABLE).
FT VARSPIC 1 98 MISSING (IN ISOFORM 2).
FT VARSPIC 99 106 EEEVIVTE -> MLGITTTT (IN ISOFORM 2).

```

FT CONFLICT 117 117 E -> K (IN REF. 2).
FT CONFLICT 142 144 NRN -> TPEI (IN REF. 2 AND 3).
FT CONFLICT 215 215 Q -> K (IN REF. 2).
FT CONFLICT 448 448 G -> E (IN REF. 2 AND 3).
FT CONFLICT 694 694 R -> G (IN REF. 2 AND 3).
FT CONFLICT 867 867 G -> S (IN REF. 2 AND 3).
FT CONFLICT 867 867 S -> A (IN REF. 2 AND 3).
FT CONFLICT 986 986 E -> EE (IN REF. 3 AND 4).
FT CONFLICT 1530 1530 V -> M (IN REF. 4).
FT CONFLICT 1581 1581 Q -> L (IN REF. 2).
FT CONFLICT 1601 1601
SQ SEQUENCE 1781 AA; 191439 MW; BA813937379FAC0F CRC64;

Query Match 5.8%; Score 122.5; DB 1; Length 1781;
Best Local Similarity 22.0%; Pred. No. 5.7;
Matches 93; Conservative 62; Mismatches 152; Indels 115; Gaps 22;

OY 40 PPTV-LVLDVSSDDEVLEVADPEVPVAPARLAPAK---PEDDSDSGAGAEAGAPAR 95
Db 192 PPTVOLLTVKKKBGAGAGAGHODPSLIGAGEASKESEPKOSTEKPEP----- 240
OY 96 TLVRRRRRLDPGEAPVVPVYSGKVOSSLNLPDSSSLKLCPSPEPD--EADLTNSGS 153
Db 241 TLKRRQSHAEISP-----PAESGA-----VECKEKGEGKQKPEPSKAE 281
OY 154 SPSEDDALPSSGP-----WRKK--LRKKCEKE---EKKMEFPDODISPLPOP 196
Db 282 SPSPVSTSETGTFKKFTQGWAGWKKTSFRKKPEKDEVEASEKKKEQPEK-----VDT 336
OY 197 SSRNKSRRKTELQKIREVNNKRLQDLRSLSPK---QHSPALQSDDDVYLVEGP----- 249
Db 337 EEDGAEVAASEKLTJASEQHPO-EPESAHEPRLSAEYKVELPS-EEGVSSGSGSEK 394
OY 250 VLPSSRLFTLTKRCRADVLRLPVRMSEPLQNVVDMANHLGVSPKILLTGESELSPT 309
Db 395 PAPLATEVEDELTVEHQEVAEVAHVHSTVEERTEEQ-----KTEVEET 437
OY 310 A--TPSTKLGVADITIDCVVLASSSEATSETSOELRLR---VQGEKHQMLETSLSPDSE 364
Db 438 ACSVPAEELVGM-----AEPQEAPEAKELVKLETCVSGEDPTQCAD--LSPDEKY 487
OY 365 -----KYLMSHYERAMGLSGHKLSFFPDGT---KLSEKELPADLG---LESGLDI 408
Db 488 LSKPPEGVSEVEMLSQERMKVQSGSPKLKLTSTGLKRLSGKKOKGKRGGDESGEHT 547
OY 409 EV 410
Db 548 QV 549

RESULT 8
TCF8_RAT STANDARD; PRT; 1109 AA.
AC 062947: 062948:
DT 20-AUG-2001 (Rel. 40; Created)
DT 20-AUG-2001 (Rel. 40; Last sequence update)
DT 20-AUG-2001 (Rel. 40; Last annotation update)
DE TRANSCRIPTION FACTOR 8 (ZINC FINGER HOMEDOMAIN ENHANCER-BINDING PROTEIN) (ZFHP).
GN TCF8.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN 11
RP SEQUENCE OF 6-1109 FROM N.A. AND ALTERNATIVE SPLICING.
RX MEDLINE=96365389; PubMed=8769566;
RA Cabanillas A.M.; Darling D.S.;
RT "Alternative splicing gives rise to two isoforms of zfhpf, a zinc finger/homeodomain protein that binds T3-response elements.";
RL DN1 Cell Biol. 15:643-651(1996).
CC -1- FUNCTION: ACTS AS A TRANSCRIPTIONAL REPRESSOR. BINDS TO E-BOX SEQUENCES IN THE IMMUNOGLOBULIN HEAVY CHAIN ENHANCER AS WELL AS IN

CC THE REGULATORY REGIONS OF MANY OTHER TISSUE-SPECIFIC GENES (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1/ZFHP-1 (SHOWN HERE) AND
CC 2/ZFHP-2; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO DELTA-EF1/ZFH-1 FAMILY OF TWO-HANDED ZINC
CC FINGER/HOMEDOMAIN PROTEINS.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use, by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: U51583; AAB17130.1; -;
DR EMBL: U51584; AAB17131.1; -;
DR HSP; P08046; 1AIG.
DR InterPro: IPR00010; Cystatin.
DR InterPro: IPR001356; Homeobox.
DR InterPro: IPR000822; Znf-C2H2.
DR Pfam: PF00031; cystatin; 1.
DR Pfam: PF00046; homeobox; 1.
DR Pfam: PF00096; zfc2h2; 7.
DR SMART; SM00389; HOX; 1.
DR SMART; SM00355; Znf_C2H2; 7.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 6.
KW Transcription regulation; DNA-binding; Nuclear protein; zinc-finger;
KW Homeobox; Repressor; Activator; Metal-binding; Repeat;
KW Alternative splicing.
FT ZN_FING 150 173 C2H2-TYPE.
FT ZN_FING 180 202 C2H2-TYPE.
FT ZN_FING 220 242 C2H2-TYPE.
FT ZN_FING 248 272 C2H2-TYPE (ATYPICAL).
FT DNA_BIND 559 618 HOMEBOX-LIKE.
FT ZN_FING 881 903 C2H2-TYPE.
FT ZN_FING 909 931 C2H2-TYPE.
FT ZN_FING 937 958 C2H2-TYPE (ATYPICAL).
FT DOMAIN 968 1109 GLU-RICH (ACIDIC).
FT VARSPIC 1 198 MISSING (IN ISOFORM 2).
SQ SEQUENCE 1109 AA; 121626 MW; BEFE291C8795DDA6 CRC64;

Query Match 5.8%; Score 121; DB 1; Length 1109;
Best Local Similarity 21.4%; Pred. No. 4;
Matches 87; Conservative 57; Mismatches 152; Indels 110; Gaps 19;

OY 61 PVEVPVAPARLAPAKPEODSDSEGAGAPARTLVRRRRRLDPGEAPVVPVY-G 119
Db 535 PAQPP---PAPATEKPESSASSAGND-----LSPQPLKMLSL 574
OY 120 KVOSSLNLPDSSSLKLCPSF--PEDA---DLTNGSSPSEDDALPSSGPMKKRLR 173
Db 575 KAYVALNAQPSTEELTKLADSVNLPLDVYKKWFKEMQAGLITGQSLPEPPSGGNTIPA 634
OY 174 KCEKEKKME-FEPDOD-----ISP-LPQSSSRNKSRRKTEALQKIREVNNKRLQ 220
Db 635 KTEEQPQPYDGNRPQEDSTRGQSPKMTNTPVLPVGSAINSRCTSSRPLNLSANRP 694
OY 221 DLRSCLSPKQHSPPALQSTDEEVYLVEGFLPOS--SRIF-----TLKIRCRADL 268
Db 695 QGVSCVSEGTQEPQVEPLDLTLPRQGGELLRRSTVSSVYQNSVQEEPLNLSGAR-- 752
OY 269 VALP-----VRMSEPLQNVVDMANHLGVSPKILLTGESELSPTAT--PSTKLGLVA 320
Db 753 -KEPKDSCVTSQEPVNVVPPSANPINIA-----IPVVAQLPTVAINDQ 798
OY 321 DLIDCVVLASSSEATSETSOELRLRVOGKEKHQML-----EISLSP--DSPKLVLN- 368
Db 799 NSVPC-----LRALANKOTILLIPQVAITYTATVSPAMQPEPPKVIQIP 841

QY 369 -SHYEAMGLSGHKLSE----FEDGTLKSGKELPADLGLSGDLIE 409
 DB 842 NGNODERODTSSGCVSEDDNDSDCTPPKKTKRAENGMYACDLCD 887

RESULT 9
 PSC_DROME STANDARD; PRT; 1603 AA.
 ID PSC_DROME
 AC P35820;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE POSTERIOR SEX COMBS PROTEIN.
 GN PSC.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_Taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92018190; PubMed=1833647;
 RA Brunk B.P., Martin E.C., Sharp E., Adler P.N.;
 RT "Drosophila genes Posterior Sex Combs and Suppressor two of zeste
 RT encode proteins with homology to the murine bmi-1 oncogene."
 RL Nature 353:351-353(1991).
 CC -1- FUNCTION: THE POLYCOMB GROUP (PC-G) GENES ARE NEEDED TO MAINTAIN
 CC EXPRESSION PATTERNS OF THE HOMOLOGIC SELECTOR GENES OF THE
 CC ANTENNAPEDIA (ANTP-C) AND BITHORAX (BX-C) COMPLEXES, AND HENCE FOR
 CC THE MAINTENANCE OF SEGMENTAL DETERMINATION.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
 CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch)
 CC -----
 DR EMBL: X59275; CAA1965.1; -
 DR PIR: S17983; S17983.
 DR FLYBASE: FBgn0005624; Psc.
 DR InterPro: IPR001841; ZnfRing.
 DR Pfam: PF00097; zf-C3HC4; 1.
 DR SMART: SM00184; RING; 1.
 DR PROSITE: PS00518; ZINC_FINGER_C3HC4; 1.
 KW Zinc-finger; Developmental protein; DNA-binding; Nuclear protein.
 FT DOMAIN 47 53
 FT DOMAIN 83 88
 FT DOMAIN 91 98
 FT DOMAIN 145 152
 FT DOMAIN 184 202
 FT ZN_FING 265 303
 FT DOMAIN 642 651
 FT DOMAIN 1066 1069
 FT DOMAIN 1185 1189
 FT DOMAIN 1214 1217
 FT DOMAIN 1391 1396
 FT DOMAIN 1458 1461
 FT DOMAIN 1517 1520
 FT SEQUENCE 1603 AA; 169999 MW; 77024F409736473 CRC64;

Query Match 5.7%; Score 119; DB 1; Length 1603;
 Best Local Similarity 18.9%; Pred. No. 8.2; Mismatches 123; Indels 114; Gaps 13;
 Matches 68; Conservative 54;

QY 45 VDLVSDSEVLEVADPEVVARLPAPAKPEDSDS---DSEGAAGPAGAPRTLVRRR 101
 DB 598 IDLSKONSVTIIDMSPEKREIVK---PLKPEKSKSKKKDKGSPKSSSSSSSSSGER 654

QY 102 RRRLLDGEAPVVPVYSGKVOSS---LNLIP-----DNSSLKLCPSEP----- 142
 DB 655 KRKSPSLVTPPLITERTIMSPGCVTLSPRYTSGAFSDPKSEFLKPALEPIKAYVE 714
 QY 143 EDEADLTNNGSSSPSEDALPSSGSPWRRKKLRKCKEKKKEEPPDDDISPLPOSSRNKS 202
 DB 715 SPERTLNINRAITPPSPVOOASAP-----KSKGNLDDSIIMKPPSCMPKSIASSKRKS 769
 QY 203 RKHTPALQKLRVKNRKLQDRSLSPKQKOSPALQSDDEVVYL-----VEG 248
 DB 770 KEPPKAVSK-----KQKLSPLPVDFEIRLPLVTNGSSGTASPKEIK 812
 QY 249 PVLPOSSR---LFTLIRCRADLVRLP-----YRMSEP-----LQNVVDHMA- 287
 DB 813 PLMPRPAPKAPMLAPRLQPAQAFAPSPPTIHNNAGVQMSAPGNRTPIAKRYQPIILKASR 872
 QY 288 -NLGVSPNRILLF-----GESLSPTAPRPT 314
 DB 873 PNPFANIPNDVNRLLKADAGTEIKSIGGSVENNSNSAQKPHLYGPKGETKMGPPALPAT 931

RESULT 10
 MSL1_DROME STANDARD; PRT; 1039 AA.
 ID MSL1_DROME
 AC P50535;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE MALE-SPECIFIC LETHAL-1 PROTEIN.
 GN MSL-1.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_Taxid=7227;
 RN [1]
 RP SEQUENCE OF 85-1039 FROM N.A.
 RX STRAIN=CANTON-S;
 RC MEDLINE=93314941; PubMed=8325488;
 RA Palmer M.J., Mergner V.A., Richman R., Manning J.E., Kuroda M.I.,
 RA Lucchesi J.C.;
 RT "The male-specific lethal-one (msl-1) gene of Drosophila melanogaster
 RT encodes a novel protein that associates with the X chromosome in
 RT males."
 RL Genetics 134:545-557(1993).
 RN [2]
 RP REVISIONS, SEQUENCE FROM N.A.
 RX MEDLINE=95300219; PubMed=7781064;
 RA Kelley R.L., Solovyeva I., Lyman L.M., Richman R., Solovyev V.,
 RA Kuroda M.I.;
 RT "Expression of msl-2 causes assembly of dosage compensation
 RT regulators on the X chromosomes and female lethality in Drosophila."
 RL Cell 81:867-877(1995).
 CC -1- FUNCTION: THE MSL PROTEINS ARE ESSENTIAL FOR ELEVATING
 CC TRANSCRIPTION OF THE SINGLE X CHROMOSOME IN THE MALE (X CHROMOSOME
 CC DOSAGE COMPENSATION). MSL-1 IS A PIONEER PROTEIN. MLE, MSL-1 AND
 CC MSL-3 ARE CO-LOCALIZED ON THE X CHROMOSOME. EACH OF THE MSL
 CC PROTEINS REQUIRES ALL THE OTHER MSLs FOR WILD-TYPE X-CHROMOSOME
 CC BINDING.
 CC -1- SUBUNIT: MSL-1 SEEMS TO FORM A TIGHT COMPLEX WITH MSL-2.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR; MSL-1 IS ASSOCIATED WITH HUNDREDS
 CC OF DISCRETE SITES ALONG THE LENGTH OF THE X CHROMOSOME IN MALES
 CC AND NOT IN FEMALES, AND IS ALSO ASSOCIATED WITH 10-20 AUTOSOMAL
 CC SITES IN MALES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch)
 CC -----


```

FT REPEAT 608 619 2-4.
FT REPEAT 623 634 2-5.
FT REPEAT 636 647 2-5.
FT REPEAT 650 661 2-7.
FT REPEAT 683 694 2-8.
FT REPEAT 717 728 2-9.
SQ SEQUENCE 872 AA; 97305 MW; C60F5B8808ED31 CRC64;

Query Match 5.5%; Score 115.5; DB 1; Length 872;
Best Local Similarity 24.5%; Pred. No. 6.4;
Matches 87; Conservative 43; Mismatches 128; Indels 97; Gaps 17;

QY 47 LVSDSDEEVLEVADPEVPAVARLPAPAKP-DQSDSDSEGAAGPAGAPRTLVRRRRRL 105
D 215 LSESHHEVYEVEDDD-----SSAKTGQKVDFDFASILLGKTRKRVRRIRK 267
QY 106 LDPGAPVVPVYSGVQSS-----LNIIPNSSLLKCPSEPPDEADLTSSGSSP 155
D 268 -----NFKSKRVKRESEHTTFODPPNLOQSSNNSSEARQDDADE-DQSDNNDSP 317
QY 156 SEDALPGSPKRLRKCEKEEKKMEFPDQDISPLPQPSRRKSKRHTEALQKLEEV 215
D 318 L-DFTLP-----MDLLKRLYKRRKN-----SGLVSSLPSEQ-----QETREE 354
QY 216 NKRLDRLRSCSP-KOHOSPALQSTDEEVVLEGPVLPQSSRLFTLTKRCRADVLVLR 274
D 355 KVLDEMDKDSLSHFQK-----IQYVDSASLPISSVFLQNGNTLPSTVNNNTVPQOLPL- 408
QY 275 MSEPLONVVDHANHL-----GVSPLRILLFGESELSPAT-----PST 314
D 409 --EPKLPATGSAHNLVEEYNOGHPNSGALQTLQPLKPTATGSANYLKMSHMQPQS 466
QY 315 LKLGAVDIIDCVLASSEATETS---QELRLRVQKEKHOMLE-ILSPDSP 364
D 467 IK-----PSTPEYTVNNGGIGQPLKPTATGSAHNLKMHISPVANPV 509

RESULT 14
TCOF_HUMAN STANDARD; PRT; 1411 AA.
ID 013428; 099408;
AC 01-NOV-1997 (Rel. 35. Created)
DT 01-NOV-1997 (Rel. 35, last sequence update)
DE 20-AUG-2001 (Rel. 40, last annotation update)
DR TRACHELE PROTEIN (TREACHER COLLINS SYNDROME PROTEIN).
GN TCOF1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96154183; PubMed=8563749;
RA Dixon J., Edwards S.J., Gladwin A.J., Dixon M.J., Lofthus S.K.,
RA Bonner C.A., Kopitvnikar K., Wasmuth J.J.;
RT "Positional cloning of a gene involved in the pathogenesis of
RT Treacher Collins syndrome. The Treacher Collins syndrome
RT Collaborative Group."
RL Nat. Genet. 12:130-136(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97250498; PubMed=9096354;
RA Wise C.A., Chiang L.C., Paznekas W.A., Sharma M., Musy M.M.,
RA Ashley J.A., Lovett M., Jabs E.W.;
RT "TCOF1 gene encodes a putative nuclear phosphoprotein that exhibits
RT mutations in Treacher Collins syndrome throughout its coding
RT region."
RL Proc. Natl. Acad. Sci. U.S.A. 94:3110-3115(1997).
RN [3]
RP VARIANTS L-439; V-810; V-1313 & G-1355, AND VARIANT TCS R-53.
RX MEDLINE=97195537; PubMed=9042910;
RA Edwards S.J., Gladwin A.J., Dixon M.J.;

```

```

RT "The mutational spectrum in Treacher Collins syndrome reveals a
RT predominance of mutations that create a premature-termination
RT codon."
RL Am. J. Hum. Genet. 60:515-524(1997).
CC -1- DISEASE: DEFECTS IN TCOF1 ARE THE CAUSE OF TREACHER COLLINS
CC SYNDROME (TCS). TCS IS A AUTOSOMAL DOMINANT DISORDER OF
CC CRANIOFACIAL DEVELOPMENT THAT OCCURS WITH AN INCIDENCE OF 1/50,000
CC LIVE BIRTHS. THE CLINICAL FEATURES OF TCS ARE BILATERALLY
CC SYMMETRICAL AND INCLUDE: (1) ABNORMALITIES OF THE EXTERNAL EARS,
CC ARESIA OF THE EXTERNAL EAR CANALS, AND MALFORMATION OF THE MIDDLE
CC EAR OSSICLES, WHICH MAY RESULT IN CONDUCTIVE HEARING LOSS; (2)
CC LATERAL DOWARD SLOPING OF PALPEBRAL FISSURES, FREQUENTLY WITH
CC COLOBOMAS OF THE LOWER EYELIDS; (3) HYPOPLASIA OF THE MANDIBLE AND
CC ZYGOMATIC COMPLEX; (4) CLEFT PALATE.
CC -----
CC This SWISS-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U40847; AAC50903.1; -;
DR EMBL; U76366; AAC51181.1; -;
DR EMBL; U84664; AAC51185.1; -;
DR EMBL; U84640; AAC51185.1; JOINED.
DR EMBL; U84641; AAC51185.1; JOINED.
DR EMBL; U84642; AAC51185.1; JOINED.
DR EMBL; U84643; AAC51185.1; JOINED.
DR EMBL; U84644; AAC51185.1; JOINED.
DR EMBL; U84645; AAC51185.1; JOINED.
DR EMBL; U84646; AAC51185.1; JOINED.
DR EMBL; U84647; AAC51185.1; JOINED.
DR EMBL; U84648; AAC51185.1; JOINED.
DR EMBL; U84649; AAC51185.1; JOINED.
DR EMBL; U84650; AAC51185.1; JOINED.
DR EMBL; U84651; AAC51185.1; JOINED.
DR EMBL; U84652; AAC51185.1; JOINED.
DR EMBL; U84653; AAC51185.1; JOINED.
DR EMBL; U84654; AAC51185.1; JOINED.
DR EMBL; U84655; AAC51185.1; JOINED.
DR EMBL; U84656; AAC51185.1; JOINED.
DR EMBL; U84657; AAC51185.1; JOINED.
DR EMBL; U84658; AAC51185.1; JOINED.
DR EMBL; U84659; AAC51185.1; JOINED.
DR EMBL; U84660; AAC51185.1; JOINED.
DR EMBL; U84661; AAC51185.1; JOINED.
DR EMBL; U84662; AAC51185.1; JOINED.
DR EMBL; U84663; AAC51185.1; JOINED.
DR MIM; 154500; -.
KW Disease mutation; Polymorphism.
FT DOMAIN 89 97
FT DOMAIN 204 207
FT DOMAIN 616 619
FT DOMAIN 919 924
FT DOMAIN 1285 1289
FT DOMAIN 1375 1386
FT DOMAIN 1398 1405
FT VARIANT 53
FT VARIANT 439
FT VARIANT 810
FT VARIANT 1313
FT VARIANT 1355
SQ SEQUENCE 1411 AA; 144312 MW; 3880203D985C2699 CRC64;

Query Match 5.5%; Score 114.5; DB 1; Length 1411;

```

Best Local Similarity 21.4%; Pred. No. 13;
Matches 84; Conservative 50; Mismatches 172; Indels 87; Gaps 16;

QY 2 AEFLRGRGPRSRG-----RGARARGRCRCRARRQSPARLLPDVIVLVSDSEEV 56
DB 715 ANPAAARASAKGTISAPGVVTAQAQORSPSKVPPV-----NPNSTV 762

QY 57 EVADPEVEP-VARLPAPAK-----PEODS-----DSDSEGAAGPAGPRTVRRRRRL 105
DB 763 LANGPASVSVGKAVATAQAQGTPEDESSSEESDESEATLQAQKPSGTHQIRAA 822

QY 106 LDP-----GEAPVVPVYSGKVOSSLNLPDSSLLKLCPSPEDEADITNSG----- 152
DB 823 LPAKESPPKGAAPTPPGKTPSAQAQKODSGSSSEESDGEAPAAVTSQVTKPPL 882

QY 153 --SSPEDDALPSGSPWR-----KKLRKCKEKKMEEFDDODISLPQ---TSSRN 200
DB 883 IFVDPNRSPAGPATPAQAQAASSTPRKARSESTANSSSESEDEVDVTPAQCTTGIRI 942

QY 201 K-----SRKHTALQKLEVNKRKLQDLRCLSPKOHSPALQSTDEVLEVEGP 249
DB 943 NYVTMTAHPRIAPKASMGASSKSSRISD-----GKQGGPATQ-----VSKNP 990

QY 250 V-LPQSSRLFTLKIRC-RADLVRLPYRMSEPLQNVVDHMANHLGVSPNRILLFGSELS 307
DB 991 ASLPLTQA--ALKVLQAQKASEAQPVARTOPSSGVDSDAVGTLPATSPQ-----STSVQ 1041

QY 308 PTAATPSTLKGVAIDIDCVVLAASSSEATFESOE 340
DB 1042 AKGTNKLKPKLPEVQAATKAPESSDSDSD 1074

RESULT 15
DYNA_RAT
ID DYNA_RAT STANDARD: PRT: 1280 AA.
AC P28023;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DYNACTIN 1 (150 KDA DYNEIN-ASSOCIATED POLYPEPTIDE) (DP-150) (DAP-150)
DE (P150-GLEU).
GN DCTN1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RX MEDLINE=91260877; PubMed=1828535;
RA Holzbaur E.L.F., Hammarback J.A., Paschal B.M., Kravitz N.G.,
RA Pfister K.K., Vallee R.B.,
RT "Homology of a 150k cytoplasmic dynein-associated polypeptide with
RT the Drosophila gene Glued.";
RL Nature 351:579-580(1991).
RN [2]
RP REVISIONS.
RA Holzbaur E.L.F., Hammarback J.A., Paschal B.M., Kravitz N.G.,
RA Pfister K.K., Vallee R.B.,
RL Submitted (DEC 1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: REQUIRED FOR THE CYTOPLASMIC DYNEIN-DRIVEN RETROGRADE
CC MOVEMENT OF VESICLES AND ORGANELLES ALONG MICROTUBULES. DYNEIN-
CC DYNACTIN INTERACTION IS A KEY COMPONENT OF THE MECHANISM OF AXONAL
CC TRANSPORT OF VESICLES AND ORGANELLES.
CC -!- SUBUNIT: LARGE MACROMOLECULAR COMPLEX OF AT LEAST 10 COMPONENTS.
CC P150(GLEU) BINDS DIRECTLY TO MICROTUBULES AND TO CYTOPLASMIC
CC DYNEIN.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: CONTAINS 1 CAP-GLY DOMAIN.
CC -!- SIMILARITY: STRONG, TO OTHER SPECIES DYNACTIN 150 KDA SUBUNIT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X61260; CAA44091.1; -
DR PIR; S16129; S16129.
DR InterPro; IPR000938; CAP-Gly.
DR Pfam; PF01302; CAP_GLY; 1.
DR PROSITE; PS00845; CAP_GLY_1; 1.
KW Motor protein; Microtubules; Dynein; Coiled coil; Cytoskeleton.
FT DOMAIN 48 90
FT DOMAIN 157 184
FT DOMAIN 214 513
FT DOMAIN 942 1048
FT DOMAIN 1184 1213
SQ SEQUENCE 1280 AA; 141929 MW; C9348CF129FAFF5C CRC64;

Query Match 5.4%; Score 113.5; DB 1; Length 1280;
Best Local Similarity 22.6%; Pred. No. 13;

Matches 87; Conservative 61; Mismatches 144; Indels 93; Gaps 17;

QY 71 APAKPPQSDSDSEGAAGPAGAPRTIVRRRRRLDPGEAPV-----PVYS 118
DB 105 SPETP---DSSASKILKREGADAAKT-----SKLRGLKPKKAPARTTTRRRPTPTAST 158

QY 119 GKVOSSLNLPDSS-SLKLICPSEPEDEAD-----LTNSGSSPEDDALPSGS 165
DB 159 GVAQPSSSLGPGSGSASAGELSSSESTPAQTPLAARILPTPALTPSGAAP---PLPSPS 214

QY 166 PMRKTLRKCKEKKMEEFPPDDISLPQSSRNKR-----KHTEALQKLEVNKRQ 220
DB 215 KEEEGLRDQYRDLKEKLE-----TLRLKRSERDAKLEKELKHKIQLQGVQEMSKMQ 266

QY 221 ---DIRSCLSPKOHSPALQSTD-----DEVVLEGGVLPQSSRLFTLKIRCRADLV 269
DB 267 EQGADQRLRLKEKEKEKALEKERYMEEMADPADIAENATLDKE-----MAEERAESE 320

QY 270 RLPEYRMSEPLQNVVDHMANHLGVSPNRILLFGSELSPTAPPS---TLKLGVAIDIDC 325
DB 321 QDEV---EALKERVDELTTDLTELKAEI-----EEKGSDGAASSYQKOLEQGNARLKDA 372

QY 326 VYVLAASSSEATPSQELRLVQKKEKHOMLEISPSPLKVLMSHYEAMGLSGHLSF 385
DB 373 LVAMRDLSSSEKQEHVKLQKLEKKNOLELVVQRQERLQOEELISQAESTI----- 422

QY 386 FDGTRKLGKELPADLGLSGDLIEV 410
DB 423 -DELK---EQVDAALGAE--EMVEM 441

Search completed: April 22, 2002, 10:09:19
Job time: 184 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 22, 2002, 10:07:05 ; Search time 26.46 seconds
(without alignments)
2277.559 Million cell updates/sec

Title: US-09-617-923-2
Perfect score: 2099
Sequence: 1 MAEPLRGGRGPRGRGARR.....GKELPADLGESGLIEYWG 412

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SPREMBL_17:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mhc:*
8: sp_mammal:*
9: sp_organelle:*
10: sp_phage:*
11: sp_plant:*
12: sp_rodent:*
13: sp_virus:*
14: sp_vertebrate:*
15: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2099	100.0	412	11 009130	009130 mus musculu
2	1982	94.4	416	11 09CVY5	09CVY5 mus musculu
3	1382	65.8	408	6 09GIZ9	09GIZ9 macaca fasc
4	154.5	7.4	1217	4 060336	060336 homo sapien
5	145.5	6.9	1072	11 035482	035482 rattus norv
6	140	6.7	1634	5 09XW25	09XW25 caenorhabd1
7	133	6.3	971	5 09XVS4	09XVS4 caenorhabd1
8	132.5	6.2	792	5 018866	018866 caenorhabd1
9	131	6.2	664	4 092541	092541 homo sapien
10	130	6.2	704	4 09H5F9	09H5F9 homo sapien
11	128.5	6.1	484	11 09CS98	09CS98 mus musculu
12	128.5	6.1	635	10 040363	040363 medicago sa
13	128.5	6.1	712	5 061708	061708 caenorhabd1
14	128.5	6.1	755	5 09V4J5	09V4J5 drosophila
15	128.5	6.1	931	13 091995	091995 xenopus lae
16	127.5	6.1	1167	5 09VNA7	09VNA7 drosophila
17	127.5	6.1	1408	5 09W0C9	09W0C9 drosophila
18	127	6.1	239	12 057114	057114 mucosa1 dis
19	127	6.1	1368	5 09V6J0	09V6J0 drosophila

20	126	6.0	856	5 021186	021186 caenorhabd1
21	125	6.0	390	5 09XWJ9	09XWJ9 caenorhabd1
22	124.5	5.9	670	4 09BW08	09BW08 homo sapien
23	124.5	5.9	757	13 09YHD2	09YHD2 gallus gall
24	124.5	5.9	827	4 09BRK5	09BRK5 homo sapien
25	124.5	5.9	879	2 09FBJ2	09FBJ2 streptomyc
26	124.5	5.9	1340	2 0911H8	0911H8 streptomyc
27	124.5	5.9	1783	4 015038	015038 homo sapien
28	124.5	5.9	1791	4 060382	060382 homo sapien
29	124.5	5.9	2286	4 09URH8	09URH8 homo sapien
30	123.5	5.9	802	4 09UG37	09UG37 homo sapien
31	123.5	5.9	815	11 09ERQ2	09ERQ2 rattus norv
32	123.5	5.9	2752	4 09UQ35	09UQ35 homo sapien
33	122.5	5.8	1152	5 09VZ23	09VZ23 drosophila
34	122	5.8	876	11 09DC22	09DC22 mus musculu
35	122	5.8	1082	5 001905	001905 caenorhabd1
36	122	5.8	1601	5 09V619	09V619 drosophila
37	121.5	5.8	678	11 099L92	099L92 mus musculu
38	121.5	5.8	1810	5 09W363	09W363 drosophila
39	121	5.8	584	11 09D672	09D672 mus musculu
40	121	5.8	886	13 070049	070049 xenopus lae
41	121	5.8	1668	4 015026	015026 homo sapien
42	121	5.8	2971	4 0915L9	0915L9 homo sapien
43	120.5	5.8	6994	5 017343	017343 caenorhabd1
44	120.5	5.7	1136	13 098UH3	098UH3 xenopus lae
45	120.5	5.7	1881	2 0917Q2	0917Q2 streptococc

ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	412 AA.
009130	009130			
AC	009130			
DT	01-JUL-1997 (Tremblrel. 04, Created)			
DT	01-JUL-1997 (Tremblrel. 04, Last sequence update)			
DT	01-JUN-2001 (Tremblrel. 17, Last annotation update)			
DE	NUCLEAR PROTEIN NIP45.			
GN	NIP45.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
NC	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=97102593; PubMed=8943202;			
RT	Hodge M., Chun H., Rengaraajan J., Alt A., Lieberman R., Glimcher L.,			
RT	"NF-AT-Driven Interleukin-4 transcription potentiated by NIP45.";			
RL	Science 274:1903-1905(1996).			
DR	EMBL: U76759; AAC52963.1; -			
DR	MGI: MGI:1329015; Nfata2.lip.			
DR	InterPro: IPR000626; Ubiquitin.			
DR	SMART: SM00213; UBO; 1.			
DR	PROSITE: PS50053; UBQUITIN_2; 1.			
KW	Nuclear protein.			
SO	SEQUENCE 412 AA; 45121 MW; DD58F5C7055C186 CRC64;			

Query Match 100.0%; Score 2099; DB 11; Length 412;
Best Local Similarity 100.0%; Pred. No. 4e-143;
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MAEPLRGGRGPRGRGARRARARAGRCPRAROSPALLIDYLVLDVSSDSEVLEVAD	60
DB	1	MAEPLRGGRGPRGRGARRARARAGRCPRAROSPALLIDYLVLDVSSDSEVLEVAD	60
QY	61	PVEVPVAPRLPAPAKPPEDSDSEGAEPAGAPRTLVRRRRRLDPGAPVVPVYSGK	120
DB	61	PVEVPVAPRLPAPAKPPEDSDSEGAEPAGAPRTLVRRRRRLDPGAPVVPVYSGK	120
QY	121	VQSSLNLIIPDNSSLLKLCPEPEDEADLTNSGSSPSEDDALPGSGPWRRKLKCKEERK	180
DB	121	VQSSLNLIIPDNSSLLKLCPEPEDEADLTNSGSSPSEDDALPGSGPWRRKLKCKEERK	180

Db 121 VOSSLNLIDNSSLLKLCSEPEDEADLTNCGSSPSEDDALPSGSPWKKLRKCKEERK 180
QY 181 KNEEFPDODISLPPOSSRNKSRKTEALQKREVNKRLODLRSCLSPKOHOSPALQSTND 240
Db 181 KNEEFPDODISLPPOSSRNKSRKTEALQKREVNKRLODLRSCLSPKOHOSPALQSTND 240
QY 241 DEVLVEGVLPPOSSRLFLTKRCRADLVRLPYRMSSEPLQNVDMANHLGVSPNRILL 300
Db 241 DEVLVEGVLPPOSSRLFLTKRCRADLVRLPYRMSSEPLQNVDMANHLGVSPNRILL 300
QY 301 FGSELSPATSTLKLGVADIIDCVVLASSSEATETSOELRLVQGEKHOEITSLSP 360
Db 301 FGSELSPATSTLKLGVADIIDCVVLASSSEATETSOELRLVQGEKHOEITSLSP 360
QY 361 DSPLKVLMSHYEAMGLSGHKLSFFPDGTRKLSGKELPADLGLESGLIEVWG 412
Db 361 DSPLKVLMSHYEAMGLSGHKLSFFPDGTRKLSGKELPADLGLESGLIEVWG 412

RESULT 2

Q9CVY5 PRELIMINARY; PRT; 416 AA.
AC 09CVY5:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE NUCLEAR FACTOR OF ACTIVATED T-CELLS, CYTOPLASMIC 2 INTERACTING PROTEIN (FRAGMENT).
GN NFATC2IP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Aikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I., Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Flisckmann W., Gaasterland T., Gissi C., King B., Kochia H., Kuehl P., Lewis S., Matsuo T., Nakai I., Pesole G., Quackenbush J., Schiml L.M., Stabili F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Kombaerts P., Norioka P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L., Wyszynski B., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S., RA Hayschizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL; AK005947; BAB24331.1; -
DR MGD; MGI:1329015; Nfatc2ip.
DR InterPro: IPR000626; Ubiquitin.
DR SMART; SM00213; UBO; 1.
FT NON_TER 1
SQ SEQUENCE 416 AA; 45551 MW; E4B46F65CC571AF5 CRC64;

Query Match 94.4%; Score 1982; DB 11; Length 416;
Best Local Similarity 97.0%; Pred. No. 1e-134;
Matches 394; Conservative 1; Mismatches 9; Indels 2; Gaps 1;
QY 9 GPSRSGRGARRA--RGARGRCPRAROSPRLIPDTVLDVLDSDSEVLEVPVPEV 66
Db 11 GTMSEVPKPRPRSRRTGANGRCPRAROSPRLIPDTVLDVLDSDSEVLEVPVPEV 70

QY 67 ARLPAPAKPEDSDSDSEGAEGPAGAPRTLVRRRRRLDPGEAVVPVYSGVQSSLN 126
Db 71 ARLPAPAKPEDSDSDSEGAEGPAGAPRTLVRRRRRLDPGEAVVPVYSGVQSSLN 130
QY 127 LIPDSSLKLCPSPEDEADLTNCGSSPSEDDALPSGSPWKKLRKCKEERKNEEFP 186
Db 131 LIPDSSLKLCPSPEDEADLTNCGSSPSEDDALPSGSPWKKLRKCKEERKNEEFP 190
QY 187 DODISLPPOSSRNKSRKTEALQKREVNKRLODLRSCLSPKOHOSPALQSTNDDEV 246
Db 191 DODISLPPOSSRNKSRKTEALQKREVNKRLODLRSCLSPKOHOSPALQSTNDDEV 250
QY 247 EGVLPPOSSRLFLTKRCRADLVRLPYRMSSEPLQNVDMANHLGVSPNRILLGSE 306
Db 251 EGVLPPOSSRLFLTKRCRADLVRLPYRMSSEPLQNVDMANHLGVSPNRILLGSE 310
QY 307 SPAPSTLKLGVADIIDCVVLASSSEATETSOELRLVQGEKHOEITSLSPSLKV 366
Db 311 SPAPSTLKLGVADIIDCVVLASSSEATETSOELRLVQGEKHOEITSLSPSLKV 370
QY 367 LMSHYEAMGLSGHKLSFFPDGTRKLSGKELPADLGLESGLIEVWG 412
Db 371 LMSHYEAMGLSGHKLSFFPDGTRKLSGKELPADLGLESGLIEVWG 416

RESULT 3

Q9GLZ9 PRELIMINARY; PRT; 408 AA.
AC 09GLZ9:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 44.6 KDA PROTEIN.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN PARIETAL LOBE;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K., Suzuki Y., Sugano S., Hashimoto K.;
RT "Isolation of full-length cDNA clones from macaque brain cDNA libraries."
RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AB050511; BAB17279.1; -
DR InterPro: IPR000626; Ubiquitin.
DR SMART; SM00213; UBO; 2.
DR PROSITE; PS50053; Ubiquitin.
KW Hypothetical protein.
SQ SEQUENCE 408 AA; 44580 MW; 5E0DD6D066AC2F24 CRC64;

Query Match 65.8%; Score 1382; DB 6; Length 408;
Best Local Similarity 70.1%; Pred. No. 1.3e-91;
Matches 293; Conservative 31; Mismatches 78; Indels 16; Gaps 5;

QY 1 MAEPLRGPRSGRGARRANGARCRAROSPRLIPDTVLDVLDSDSEVLEVPV 58
Db 1 MAEPLRGPRSGRGARRANGARCRAROSPRLIPDTVLDVLDSDSEVLEVPV 52
QY 59 ---ADPVEVPVAPRLPAPAKPEDSDSDSEGAEGPAGAPRTLVRRRRRLDPGEAV 114
Db 53 ARCAADEVEVAPSEPPGVAASRDSDSDSEGAADARPPREVRRRRLVLDPEGAPLV 112
QY 115 PYSGVQSSLNIPDSSLKLCPSPEDEADLTNCGSSPSEDDALPSGSPWKKLRK 174
Db 113 PYSGVQSSLNIPDSSLKLCPSPEDEADLTNCGSSPSEDDALPSGSPWKKLRK 171
QY 175 CEKEEKKNEEFPDODISLPPOSSRNKSRKTEALQKREVNKRLODLRSCLSPKOHOSP 234
Db 172 -DKEEKKTEIIDLNDNSPLSPSPSRKSRKTRAKKLSEVNKRLODLRSCLSPKOHOSP 230


```

OY      235 ALOSTDEVLVVEGVGFLPOSSRLFFELKRCRADIVRLPVHRSPELOWNDHMNHGVSP    296
           | :|:||||||| ||: ||| |||||||||:|||||||:|||||||:|||||||
Db       231 BOOGODEVVLVEGGFTLETRPLRPLKRADLVRLPRLRSEPLQSVDMHMHGVSP    290
           | :|:||||||| ||: ||| |||||||||:|||||||:|||||||:|||||||
OY      295 NRILLFGESELSPATPFPSTLKLGVADIIDCVVLASSSEATEFSQEELRLRVQGEKHOML    354
           |:|||||||:|||||||:|||||||:|||||||:|||||||:|||||||
Db       291 SKRIILFGETEISLPATPFRLLKLGVAIDIICVLASSPEATETSRQLQVRQGEKHQTL    350
           | :|:||||||| ||: ||| |||||||||:|||||||:|||||||:|||||||
OY      355 ELSLSPDSPLAVLMASHYEAMGLSGHKLSFFEDOTIKLSGKEPLADLCESGDILEWVG    412
           | :|:||||||| ||: ||| |||||||||:|||||||:|||||||:|||||||
Db       351 EVSLSRDSPLTMTMSHYEAMGLSGRKLSFFPDGTKLSGRELPAIDMGESGDILEWVG    408
           | :|:||||||| ||: ||| |||||||||:|||||||:|||||||:|||||||

RESULT      4
060336
ID      060336          PRELIMINARY;          PRT;   1217 AA.
AC      060336;
DT      01-AUG-1998 (TREMBLrel_07, Created)
DT      01-AUG-1998 (TREMBLrel_07, Last sequence update)
DT      01-JUN-2001 (TREMBLrel_17, Last annotation update)
DE      KIAA0596 PROTEIN (FRAGMENT).
GN      KIAA0596.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
CC      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE-BRAIN;
RX      MEDLINE=98290545; PubMed=9628581;
RA      Nagase T., Ishikawa K., Miyajima N., Tanaka A., Kotani H., Nomura N.,
RA      Ohara O.;
RT      "Prediction of the coding sequences of unidentified human genes. IX.
RT      The complete sequences of 100 new cDNA clones from brain which can
RT      code for large proteins in vitro.";
RL      DNA Res. 5:31-39(1998);
DR      EMBL; AB011168; BAA25522.1; -.
DR      InterPro; IPR002114; PTS_HPr_ser.
DR      InterPro; IPR001680; WD40.
DR      Pfam; PF00400; WD40; 7.
DR      PRINTS; PR00320; GRPROTEINBRPT.
DR      SMART; SMU0320; WD40; 7.
DR      PROSITE; PS00589; PTS_HPR_SER; UNKNOWN_1.
DR      PROSITE; PSS0082; WD_REPEATS_2; 1.
DR      PROSITE; PSS0294; WD_REPEATS_REGION; 1.
KW      Repeat; WD repeat.
FT      NON_TER      1
SQ      SEQUENCE      1217 AA; 131097 MW; 0BC4EAC6722BEF5 CRC64;
```

	Query Match	7.4%	Score 154.5	DB 4	Length 1217
	Best Local Similarity	23.5%	Pred. No. 0.0085		
	Matches 103	Conservative 48	Mismatches 164	Indels 129	Gaps
Qy	6 RGRGPRSKCGAGARARGACRCRPARQSPARLLPTDVLVDIVSDSEVELEVADPEVP	65	: : : : : : :		
Db	455 RRGKGKQGQPSSPPRAGS-----PNRRQAASMLSPGA-----LSDSKKEEDECTEEELP	506	: : : : : : :		
Qy	66 -----VALRPAKPEODSDSDEGAEE-----PGAPARTLVRRRRRL	106	: : : : : : :		
Db	507 ALPYLAKSTKKALSVSPALPRSLSHMKSRAQESVGLFDPAAPAANBGP---RRRGWV	563	: : : : : : :		
Qy	107 DPGCAPVVYVSGKYQSILNIPNSSLLKLCP--EP-EDBADLNISGSSPEDDALPS	163	: : : : : : :		
Db	564 QRP-----VELSVSRMDLRQLETLAPSLOPQSDSLAIIPSGRGKHGOEALET	612	: : : : : : :		
Qy	164 G-----SPWRKTLRKCC-----EKEX-----KKGEFPPODISPLPQPS---SRNK	201	: : : : : : :		
Db	613 SLTSONEKPFRPASQPCSYPIHRIIRLLSQEBGVAAQLPEAFIEDGLVIYFEPSPDNPMOT	672	: : : : : : :		
Qy	202 SRKHTAEALQR-----IREVNKRLLQD-----LRSLCLSPKHQNSPALQSTDEEVYL	245	: : : : : : :		

```

Db      673  SEIYQAAKARTLGRVYPGSSSEKHPDAGCAVDYSSCLSPHE--PIEDSESTEPLS 730
OY      246  VEGPVLPOSSRLFTLKIRCADLRLPRMSEPLQNVVDHANHLGVSPNRILLFGESE 305
Db      731  VDG-----ISDLLEEPAGDEEEEBEGGMP-----YGLQE 762
OY      306  LSPRTATP-----STLKLGVADIIDCVLLASSEATETSOELRLRQGEKKNOMLE 355
Db      763  GSP-OTPEQEOFLKHNETTLASGAAPGAP-VQYPERSESRSSIRFLQVOTRPLREP-- 818
OY      356  ISLSPDSPKLKVLMSHYEAMGLSGHK 381
Db      819  ---SPSSSLALMSRPAPVQASSEQ 841

```

```

RESULT      5
035482
ID      035482      PRELIMINARY;      PRT;      1072 AA.
AC      035482;
DT      01-JAN-1998 (TREMblrel. 05, Created)
DT      01-JAN-1998 (TREMblrel. 05, last sequence update)
DT      01-JUN-2001 (TREMblrel. 17, last annotation update)
DE      HIGH MOLECULAR-WEIGHT NEUROFILAMENT.
GN      NF-H.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_Taxid=10116;
[1]
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=91038277; Pubmed=2230956;
RA      Chin S.S., Liem R.K.;
RT      "Isolated rat high-molecular-weight neurofilament (NF-H)
RT      coassembles with vimentin in a predominantly nonphosphorylated form."
RL      J. Neurosci. 10:3714-3726(1990).
[2]
RN      [2]
RP      SEQUENCE FROM N.A.
RA      Chin S.S.M., Liem R.K.H.;
RL      Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
CC      -1. SIMILARITY: TO THE IMMEDIATE FILAMENT FAMILY.
DR      EMBL; AF031879; AAB87068.1; -.
DR      InterPro; IPR001664; IF.
DR      Pfam; PF00038; filament; 1.
DR      ProSite; PS00226; IF; 1.
KW      Coiled coil; Intermediate filament.
SEQUENCE      1072 AA: 115348 MW: 899A146D457A4D78B CRC64:

```

Query Match	6.9%	Score 145.5	DB 11	Length 1072
Best Local Similarity	22.1%	Fred. No. 0.032		
Matches	95	Conservative	63	Mismatches 166; Indels 105; Gaps 20.
QY	4	PLRGGRPSRGGRRGARRRGARGRCPPRARRQSPARL-IPDTVLVDLVSDEEY-----	55	
Db	613	PVEAKSPAEKASPAVSKSFC-EAKSPAEKASPAEYKSPATVKSPEVEAKSPAEEKSPVYWK	671	
QY	56	--LEVADVENVPARLPAPAKPEQSDSDSEGAAGCPAGAPRTVLRRRRRRLDPEGAPV	113	
Db	672	SPAEEKSPVEV-----KSPASVKSPEAKSPAGAKSPAAKSPVYAKSPAEEKSPAERK-	725	
QY	114	VPVYSGKVQSSINLLIPDNSSLLIKCPSEPEDEADLTN-----SGSSPSSEDA--LP	162	
Db	726	-PPAEKSPAEEKSPAEEKS-----PAEAKSPAEEKSPVEVKSPEKASPVKEGAKSLAE	779	
QY	163	SGSPRRKRLRKCKCEEKEMEEFPDODISPLQPFS-SRKRSR-----KITEALQKLEVWK	217	
Db	780	AKSPEKASPVYKEELTKPAEYKSPPEKASPMKEEAKSPKAKTLDVYKSPKATTPAKEEAK	839	
QY	218	RLDRLRCLSTPKOHQSPALQSDTDEVEVLVEGVLVLOOSSSLFTLKIRCRADLRLP--AR	274	
Db	840	RPADIR---SPQVYKSPAEEKS-----PEKETRTKEKAPPKKEEYKSPVEEYK	886	
QY	275	MSEPLQNVVDHANLGYSPNRIILLFGESLSLPATPSTLKGVAADIIDCVLASSSEA	334	

```
Db 887 AKEP-----PKRV-----EEKTP-ATPKTE-----VKESKDDA 915
Oy 335 TETSOELRLRVOGKEKHOMLEISLSPDPLKYLMSHYEEMAMLSGKLSFFDGTKLSGK 394
Db 916 PKEAO-----KKPAEEKEPLTE--KKRDSRGEAKKEEAKKKAAPR-----E 956
Oy 395 ELPADIGLE 403
Db 957 EPPAKLGVK 965

RESULT 6
O9XW25 PRELIMINARY; PRT; 1634 AA.
AC O9XW25;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Y18D10A.1 PROTEIN.
GN Y18D10A.1
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nemtoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN NM
RP SEQUENCE FROM N.A.
RA Harris B.;
RA Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Lalster N., Latrelle P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Koopa A., Saunders D., Shownkeen R.,
RA Smaidon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; AL034393; CAA22308.1; -.
DR InterPro; IPR000637; AT_hook.
DR PROSITE; PS00354; HMGL.Y; UNKNOWN_1.
DR SMART; SM00384; AT_hook; 1.
SQ SEQUENCE 1634 AA; 179058 MW; DC20372F3AAAF48D0 CRC64;

Query Match 6.7%; Score 140; DB 5; Length 1634;
Best Local Similarity 22.0%; Pred. No. 0.13;
Matches 88; Conservative 54; Mismatches 152; Indels 106; Gaps 18;
```

```
Db 643 RLBDTAKTATVYHPGPPLTRKMERKAPPAVTSSKKEPKNAGSDSSINEEHEDE 702
Oy 243 VLVAGPV--LPQSHRLFTLIKIRCRADLVRLVPRMSEPLQNVDMANHLGVSPRIILL 300
Db 703 TMILEEQILDLPGQPSQGEPRISGSEL-----LDGEFDSEHSGTVPS----- 747
Oy 301 FGESELSPTAPSTYKLGVAIDIICVLASSEATETSQE 340
Db 748 --APELTKNPAPPV-----PEASEASAE 768

RESULT 7
O9XVS4 PRELIMINARY; PRT; 971 AA.
AC O9XVS4;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE C25A1.10 PROTEIN.
GN C25A1.10
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nemtoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN NM
RP SEQUENCE FROM N.A.
RA Mortimore B.;
RA Submitted (OCT-1996) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Lalster N., Latrelle P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Koopa A., Saunders D., Shownkeen R.,
RA Smaidon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; Z81038; CAB02755.1; -.
DR InterPro; IPR002965; P-rich_extensn.
DR PRINTS; PRO1217; PRICHEXTENS.
SQ SEQUENCE 971 AA; 101096 MW; 832936115B58709B CRC64;

Query Match 6.3%; Score 133; DB 5; Length 971;
Best Local Similarity 23.3%; Pred. No. 0.23;
Matches 60; Conservative 35; Mismatches 115; Indels 48; Gaps 8;
```

DB 466 SDSDEKPAKSTPAKI 483

RESULT 8
ID 018866 PRELIMINARY: PRT: 792 AA.

AC 018866;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE COSMID C55C3.
GN C55C3.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Barks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkhen R.,
RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Woessne J., Stellyes L.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U53335; AAA96170.1; -
SQ SEQUENCE 792 AA; 87338 MW; 32F6B61C3548FA28 CRC64;

Query Match 6.3%; Score 132.5; DB 5; Length 792;
Best Local Similarity 20.9%; Pred. No. 0.19;
Matches 90; Conservative 50; Mismatches 135; Indels 155; Gaps 19;

QY 11 RSGGCGARAGRCRCRARGSPARLPDVIYVLDVSDSDSELEVAADPEVAPARLP 70
DB 309 RRRGGQIKKKKPAK--PLKRTTP-----NNSDEKLKKMSP----- 343
QY 71 APAKPEODSDSD-----SEGAEGPAGAPRTLVRRRRR-----LLDPG 109
DB 344 -PKKPLQATDSIDLPPLPNNMATVPESEKSNRSPQSKINIVRNGGTVNTDLKDS 402
QY 110 EAPVVPVYSGKVQSSLLNLPDSSSLIKLCPSEPEDEADLTNSGSSPEDEDALPGSPMK 169
DB 403 EARAIP-----KC-DKLTDFQTSTNSEQSPA--DAISAVAP-TK 437
QY 170 KLRKCKEKKKEEPPD-----ODI-----SPLPQPSRKNK----- 201
DB 438 LSGSQSQSRQETEKMPNFRLEQDSKFKPAEAPLPKPEERKKGSTQSTTEPTKLASQ 497
QY 202 SKRHTALQKLEBVKRLQD--LRSLSPKQHQSPALQSTDD----- 241
DB 498 SPSEKQEIQMPNFRLEQDSKFKPAEAPLPKPEERKKGSTQSTTEPTKLASQ 557
QY 242 -----EYVLVEGVPVLPQSSRLFTLKIRCADLVLPVPMSEPLQNVYDH-----MAN 288

DB 558 TISVPPTKLVGTQSPSEKQEIQMPNFRLEQDSKFKPAEAPLPKPEERKKGSTQSTTEPTKLASQ 617
QY 289 HLGVSPPNRILLFGESLSPATPSTLK-----IGVADIIDCVLA-----SS 331
DB 618 PLTIS-----RVAFG-SPIAKPPRSPLOAPLETLPPTIDAPTAIETASAEFSS 672
QY 332 SEATETSOEL 341
DB 673 SHSMDPSNSL 682

RESULT 9
ID 092541 PRELIMINARY: PRT: 664 AA.

AC 092541;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE MYELOBLAST KIAA0252 (FRAGMENT).
GN KIAA0252.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BONE MARROW;
RX MEDLINE=97191544; PubMed=9039502;
RA Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayashi Y., Ohara O.,
RA Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. VI.
RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT analysis of cDNA clones from cell line KC-1 and brain.";
RL DNA Res. 3:321-329(1996).
DR EMBL: D87440; BAAT3382.1; -
FT NON_TER
SQ SEQUENCE 664 AA; 75809 MW; 2C6537A98BFBADB CRC64;

Query Match 6.2%; Score 131; DB 4; Length 664;
Best Local Similarity 21.5%; Pred. No. 0.2; Indels 148; Gaps 19;
Matches 101; Conservative 69; Mismatches 151; Indels 148; Gaps 19;

QY 37 RLIPDVIYVLDVSDS--DEVELEVA-----DPVAVPARLPAPAKPEODSDSDSEGA 87
DB 2 RVIYDSDTSDSGSDNDLDELISLAKRKSDSEKPEPVSQ-PAASDSETSDDDEMT- 59
QY 88 EGPAGAPRTLVRRRRRLDPGEAPVVPYSGKVQSSLLNLPDSSSLIKLCPSEPEDEAD 147
DB 60 ---FGSNKKKKKKKARKIEKG--TMKKQANKTASSGSSDKD--SSAESAPBEGEVSD 112
QY 148 LTNSSGSSPEDEDALPGSGFW----- 167
DB 113 DSNSSSSSSSDSDSEDEFGDGLMGDEEDRARLEQMETEKEREQELFNRIERREYL 172
QY 168 -----KKILR--KKCEKEKKKEEPPDODISPL-----PQPSRKRKR-----RH 205
DB 173 KRFEELKKLTKAKKKEKKKKKEEKKKQEEBQKKLTQIQESQVTSNKKRRSKDEKDLKK 232
QY 206 TEALQRLR-EVNRRLQDLRSLCLSPKQ-----HOSPALQSTD 240
DB 233 SQAMEELKAERERKKRNTALLAKQPLKTSSEYSDDEEEDDKSSKSDRSRTSSSD 292
QY 241 DEYVLVEGVYVQSSSLFLTKIRCADLVLPVPMSEPLQNVYDHANHLGVSPNILL 300
DB 293 EEEKEKEIP--PKSQ-----PVSLPEEL-----NVRRLRHLERW 326
QY 301 FGSEELSPATPSTLKLG-----VADIIDCVLIASSSE--ATETSOELRLRYOG 347
DB 327 CHNPFPAKTVTGCFAVIGIGNHNKSPYVAETGTGVTETAKYQLQGTGRTNKGLOLRGN 386
QY 348 KEKHQMLEISLSPDPSLKVLMSHYEFAMGILSGHKLSFFDGTGLSKREL 396

Db 387 DORFERLEFVSNOEFTSEEMK-WKEAMFSAGMOLPTL---DEINKKEL 431

RESULT 10

OC 09H5F9 PRELIMINARY: PRT: 704 AA.

ID 09H5F9

AC 09H5F9

DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DE CDNA: FLJ23471 FIS, CLOVE HS11969.

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RP [1]

RP SEQUENCE FROM N.A.

RC TISSUE=HUMAN SMALL INTESTINE;

RA Kavadaba A., Hiki J. T., Kodatake N., Inagaki H., Ikema Y., Okamoto S.,

RA Okitani R., Ota T., Suzuki Y., Obyashi M., Nishi T., Shibahara T.,

RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;

RT "NEO human cDNA sequencing project";

RL Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.

DR EMBL: AK071124; BAB15667.1; -

DR InterPro: IPR002965; P_Rich_extensn.

DR PRINTS: PRO1217; PRICHEXTENS.

DR SEQUENCE 704 AA; 75768 MW; 858F94E8A2C1F8C6 CRC64;

Query Match 6.2%; Score 130; DB 4; Length 704;

Best Local Similarity 22.6%; Pred.No.0.25;

Matches 100; Conservative 63; Mismatches 173; Indels 106; Gaps 20;

QY 10 PPSRCRGARARARGARCGRCPRAROSPARLIPTVLVDVSDSEDEYLEVADPV----- 62

Db 286 PKTEAPQASPLAKPLQSSSPRVGLGSPMEPPAPL---STSTGASALPPAGRRNLAE 341

QY 63 -----EVYARLPAPAKPPQSDSDSEGAEGPAG-----AP 94

Db 342 SSGVGRVAGASRPKEAPPAKGGKSTLLTQDMSTSLQEGQEDGPAGWRANLKEVDRKSPAE 401

QY 95 RFLVRRRRRLDP--GEAP--VVVYSGKVQSSSLNI--PDNS-----SLTKCPSE 141

Db 402 KFLKKEPPALAEPRAGEAPRVSGSFASVHITLTPVPRDTPPPASGSPSLPARSPSP 461

QY 142 PEDEADLTNSGSSPSEDDALPSGSPWKKRLRKCKEKK--MEEPDODISP-----LP 194

Db 462 PR-RRRLAVASLDVCDNMLRPEPPQGEARVQSKWEKKPHLQGRPLSPANVPALP 520

QY 195 QSSSRKSKRHTEALQKLEVNKRLODLRSCISPKOHOSPALQST-----DD--EVVLY 246

Db 521 GETVSPVRLHPDYLSLSP-DEIQROLDIERRLDALELRVELEKRLRAAEGDAEDSLAV 579

QY 247 EGYPVLPQSSRLFTLKIRCADLY-RLPVRMSEPLQNVVDHMANHGVSPNRILL---FG 302

Db 580 DMFWLHIEKQL---LRQSELMYKSKAORLEQOQIDIGELRRLMAKPAKLSIQERRR 636

QY 303 ESELSPATPSTLKGVADIIDCVVLASSSEATEETSOELRLRVQGEKHOMLAISLSPDS 362

Db 637 EDEL-----LEQVYVTVND-----RSDIVLSDLEDRLRQ--EEDQMLRDM----- 676

QY 363 PLKVLMSHYEAMGSGKHLSP 384

Db 677 -----ERLGIQRRKSKF 688

RESULT 11

OC 09CS98 PRELIMINARY: PRT: 484 AA.

AC 09CS98

DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE 3930402D05RIK PROTEIN (FRAGMENT).

GN 3930402D05RIK.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RP [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=EMBRYO;

RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi T., Fukuda S.,

RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamada K.,

RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,

RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,

RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,

RA Schriml L.M., Steadil F., Suzuki R., Tomita M., Wagner L., Washio T.,

RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barns G.,

RA Blake J., Boitelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,

RA Guenichon S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,

RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,

RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,

RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitlaker C., Wilming L.,

RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Khtsuki S.,

RA Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection.";

RL Nature 409:685-690(2001).

DR EMBL: AK014457; BAB29364.1; -

DR MGI: MGI:1921269; 3930402D05RIK.

FT NON-TER 1

FT SEQUENCE 484 AA; 55413 MW; 328371FE52141F54 CRC64;

Query Match 6.1%; Score 128.5; DB 11; Length 484;

Best Local Similarity 20.6%; Pred.No.0.2;

Matches 89; Conservative 63; Mismatches 158; Indels 121; Gaps 17;

QY 2 APLRGRGPRSGRGARARGA-----RGRCPAROSP----- 35

Db 103 AERDERDRRSGEGKEKEKERVAKDRDRDKGRERKRSKNGEHTDPPREKSRQADREK 162

QY 36 -----ARLIPTVLVDVSDSEYLEVADPVYVARLPAPAKP-----BQD 78

Db 163 KSSSGEISRLKLDGSPFDVKAMEAD-----ISVGASR--SFLKRSKRKSHSLBGD 214

QY 79 SPSDSEGAAGAGAPRLVRRR-----RRLDPGAPVVPVYSGKVQSSSLNI 128

Db 215 SPSDAEVEA-GPAGODEPEVMENAEVPSLRLRPPGSARAPAPRVKROESTETLY 273

QY 129 PNSSSLKLCPEPEDEADLTNSGSSPSEDDA--LPSSPWWKKRLRKCKEKKMEFP 186

Db 274 VDRSSSGKTYSV-----IIDSQNSDNEDDEQFYVEAP-----QLSEIA 313

QY 187 DODISPLPQSSRNKSRKHTALQRLREVNKRLODLRCLSP-KOHOSPALQST---DDE 242

Db 314 DIDWVP-----SGELEDEKKGGLVKILETKKDYKLOQSLKPGEKERSLFESAMKKKED 370

QY 243 VVLVEGYPVLPQSSRLFTLKIRCADLYRLPVRMSEPLQNVVDHMANHGVSPNRILLFG 302

Db 371 IVSKIEIKLRVS-----IOTLCKSAL-----PLCKINDYIOEDVDANQNLQMLWS 416

QY 303 ESELSPATPSTLKGVADIIDCVVLASSSEATEETSOELRLR-----VQGEKXH 351

Db 417 ENRQHAHALSQE-----OSTIDSAYEPLKAEISELEQOIRDOQDKICAVKANILKNEKI 471

QY 352 QMLEISLSPDS 362

Db 472 QKAVHSINLSS 482


```
AC Q9V4J5;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE CG12165 PROTEIN.
GN CG12165.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Sulton R.C., Rogers Y.-H.C., Blake J.R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Abmayyan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Bens P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bock J., Brockstein F., Brotlier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu L.B., Davies P.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durkin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hosten D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaisl M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasro P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puti V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spirdling A.C., Stapleton M., Strong R., Sun E.,
RA Svitsks R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodard T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003841; AAF59275.1; -.
DR Flybase: FBgn0033156; CG12165.
SQ SEQUENCE 755 AA; 83536 MW; 5DA71535FF3EECD1 CRC64;
```

```
Query Match 6.1%; Score 128.5; DB 5; Length 755;
Best Local Similarity 24.1%; Pred. No. 0.35;
Matches 77; Conservative 48; Mismatches 122; Indels 73; Gaps 16;
```

```
QY 71 APAKPEQSDSDSEGAAGPAGPRT-LVRRRRRL-----DPGEAPVVPVYSGKVQ 122
   ||| ||| : : : |||
DB 41 AEAKPR-----DSAPAKSQENSSVTPQOTKKRPKRLLTSLAEDQNPPEADATANTTSAR 95
   ||| ||| : : : |||
QY 123 SSANLPDSSSLKLCPSPEDEADLTNSSSSP-SEDDALPSSGSPR-KLTKKCKEK 180
   ||| ||| : : : |||
DB 96 QSTRV--SNSQLALAEDEHNSTASLPPPVVSADTTTSGRPRAKL-----KTEK 148
   ||| ||| : : : |||
QY 181 KMEFPDQDISPLPPSSSRK-RHTEALQKL-R-EVNRKLODLRSC-----LSP 228
   ||| ||| : : : |||
DB 149 LKE-----PSLNKKMRPSPSEELVYKVESEQVQFNSTSQALAEKKLAE 197
   ||| ||| : : : |||
```

```
QY 229 KOHSPALQSTDEEVVLEGGPYLPQSSRLFTLKIRCADVLRLPYRMSEPLQNVHMAN 288
   ||| ||| : : : |||
DB 198 PELPEPAETAEQKRPPEASVTEDEVNTTKTLRVYKRE--KLSTEAVPPLTNVAVS-TAN 254
   ||| ||| : : : |||
QY 289 HLGVPENRLLLFGESELSPTATPSTKLGVADITDQVYLAASSSEATTSOELRLRVGK 348
   ||| ||| : : : |||
DB 255 VTTVS-----SVTTEARP-----DDTVASNTTSTTEVSKVKRRKKDV 293
   ||| ||| : : : |||
QY 349 EKQMLEI-----SLSPDSPL 364
   ||| ||| : : : |||
DB 294 ESHRPIKVERESDLKSSPV 313
   ||| ||| : : : |||
```

```
RESULT 15
QY 091995 PRELIMINARY; PRT; 931 AA.
AC Q91995;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CALPASPATIN.
GN CALP3.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=83355;
RN [1]
RP SEQUENCE FROM N.A.
RA Mairacchi S., Rossi C., Nardi I.;
RT "Identification of different forms of calpastatin mRNA co-expressed in
RT the notochord of Xenopus embryos.";
RL Submitted (Feb-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AJ271209; CAB71173.2; -.
DR InterPro: IPR001259; Calpain_inh1b.
DR Pfam: PF00748; Calpain_inh1b; 5.
SQ SEQUENCE 931 AA; 96825 MW; D71A3526253299E CRC64;
```

```
Query Match 6.1%; Score 128.5; DB 13; Length 931;
Best Local Similarity 21.8%; Pred. No. 0.45;
Matches 83; Conservative 60; Mismatches 150; Indels 87; Gaps 18;
```

```
QY 67 ARLPAPA-KPEQSDSDSEGAAGPAGPRTLVRRRRRLDPGEAP-----V 113
   ||| ||| : : : |||
DB 32 SKTPAPSGKTMNPVPTQPKGATGAPAAATK-----PTAAPSALAKTTAAPSA 79
   ||| ||| : : : |||
QY 114 VVYVSGKVQSSNLIPDSSSLKLCPSPEDEADLTNSSSPSEDDALPSSGPMRKKLRK 173
   ||| ||| : : : |||
DB 80 VPAAATKKPTAA--PSTAPVAVVTKPTETKPSA-----GKSPKQDTKP--TPGKQKPD 128
   ||| ||| : : : |||
QY 174 KCEKEKKMEEPDDDISPLPQ-PSSRNK-----SKRHTEALQKREVNRK 218
   ||| ||| : : : |||
DB 129 K-SKEEKKSDQKATAPVQ-PKVTSPSAGKVSAAVATGFAAAGATAAATVAVGSTQKDES 186
   ||| ||| : : : |||
QY 219 LQDLNSCLSPK--QHOSPALQSTD--DEVVLVEGPLYPQSSRLFTLKIRCADVLRLPV 273
   ||| ||| : : : |||
DB 187 KEERKAVIDQKVSASAPAKTGADTALDELDTLG-----SPADIPESPK 231
   ||| ||| : : : |||
QY 274 RMSEPLQ--VYDHMANHNG-----VSPNRILLGSELSPTATPSTKLGVADITICV 326
   ||| ||| : : : |||
DB 232 FPGPEIQDTVVVSKVVEELGKRDRHTTPVYRKLLQKGGKMAPPTPPLVAEASMDDDLIV 291
   ||| ||| : : : |||
QY 327 -VLASSSEATTSOELRLRVQK--EKHOMLEISLSPDSPVLVMSHYEAMGLSGHK- 381
   ||| ||| : : : |||
DB 292 AALSSGFKSSQCPYKKRKLKEEKLEKPKSANTVATQPOLQKDKVKAATAEFSKASIQK 351
   ||| ||| : : : |||
QY 382 --LSFFDGTKLSGKELPAD 399
   ||| ||| : : : |||
DB 352 EVKATPAEASKOSTQAIQPD 371
   ||| ||| : : : |||
```

Search completed: April 22, 2002, 10:11:59
job time: 294 sec

This Page Blank (us30)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 22, 2002, 10:06:15 ; Search time 22.6 Seconds

(Without alignments)
1350.363 Million cell updates/sec

Title: US-09-617-923-2

Perfect score: 2099

Sequence: 1 MAEPLRGGRSGRGAR.....GKELPADLGESGLIEVWG 412

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Maximum Match 100%

Listing first 45 summaries

Database :
A.Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2099	100.0	412	18	AAW34091 Mouse NF-AT intera
2	2099	100.0	412	20	AAW08331 Mouse NIP45 protei
3	1399.5	66.7	419	20	AAW08330 Human NIP45 protei
4	623	29.7	138	22	AAW94289 Human protein sequ
5	132	6.3	586	18	AAW10423 Cytomegalovirus UL
6	126.5	6.0	676	22	AAW64373 Human hepatome cel
7	124.5	5.9	518	21	AAW53322 Human colon cancer
8	124.5	5.9	555	22	AAW57073 Human colon cancer
9	124.5	5.9	671	21	AAW99426 Human PRO1604 (UNQ
10	124.5	5.9	671	22	AAW66175 Protein of the inv
11	123	5.9	950	20	AAW33298 Human membrane spa

12	122.5	5.8	1780	19	AAW53863 Human gravin polyp
13	122.5	5.8	1780	21	AAW15380 Human gravin prote
14	121	5.8	2971	21	AAW41231 Human ORF995
15	121	5.8	2972	22	AAW50363 Human SRCAP. Homo
16	121	5.8	3118	22	AAW50362 Human SRCAP. Homo
17	120.5	5.7	565	19	AAW61247 Streptococcus pneu
18	120.5	5.7	1881	21	AAW44506 Streptococcus pneu
19	120	5.7	669	19	AAW37483 Mouse liver cancer
20	118.5	5.6	508	21	AAW77945 A. thaliana enviro
21	118.5	5.6	846	21	AAW71057 Human membrane tra
22	118	5.6	576	20	AAW49541 Human PRO201 (Nsp1
23	118	5.6	576	20	AAW49545 Human Nsp1 mutant
24	118	5.6	576	20	AAW49546 Human Nsp1 mutant
25	118	5.6	576	20	AAW06477 Human tumour-assoc
26	118	5.6	576	21	AAW93684 Amino acid sequenc
27	118	5.6	576	21	AAW51932 Human PRO201 prote
28	118	5.6	576	21	AAW51939 Human PRO201 prote
29	118	5.6	987	22	AAW84633 Human protein sequ
30	117.5	5.6	815	22	AAW25561 Human CENP-C anti
31	117	5.6	979	14	AAW34783 Human tumour suppr
32	117	5.6	2663	22	AAW98612 Human Actinus L pro
33	115.5	5.5	1341	21	AAW85657 Human Nsp1 mutant
34	115	5.5	576	20	AAW49547 Peptide #3033 enco
35	114.5	5.5	92	22	AAW16599 Peptide #3120 enco
36	114.5	5.5	92	22	AAW29083 Peptide #2994 enco
37	114.5	5.5	92	22	AAW04312 Human protein sequ
38	114.5	5.5	454	22	AAW95394 Human striated mus
39	113.5	5.4	661	19	AAW77048 Human cell signal
40	113.5	5.4	661	21	AAW70078 Human cell signal
41	113	5.4	519	21	AAW44247 Human normal bladd
42	113	5.4	605	22	AAW12251 Human normal bladd
43	113	5.4	783	20	AAW60344 Ubiquitin-like dom
44	112.5	5.4	95	20	AAW87985 Maize protein enco
45	112.5	5.4	610	12	AAW10923

ALIGNMENTS

RESULT	1
AAW34091	standard; Protein: 412 AA.
ID	AAW34091.
AC	AAW34091.
XX	
DF	18-MAY-1998 (first entry)
XX	
DE	Mouse NF-AT interacting protein 45.
XX	
KW	NF-AT Interacting Protein 45; NIP45; yeast two-hybrid assay; mouse;
KW	Rel homology domain; RHD; T cell; transcription factor; cancer;
KW	Interleukin-4; IL-4; development; Th1; Th2; cytokine; allergy;
KW	autoimmune disease; transplantation.
XX	
OS	Mus sp.
XX	
FT	Key
FT	Region
PN	Location/Qualifiers
PN	6..37 /note= "highly basic region"
XX	
PD	30-OCT-1997.
XX	
PF	23-APR-1997; 97MO-US06708.
XX	
PR	25-NOV-1996; 96US-0755592.
PR	23-APR-1996; 96US-0636602.
XX	
XX	25-NOV-1996; 96US-0755584.
PA	(HARD) HARVARD COLLEGE.
XX	
PI	Glimcher LH, Ho I, Hodge MR;

```
XX WPI, 1997-535556/49.
DR N-PSDB; AAT93045.
XX
XX Production of cytokine(s) associated with Th2-type helper T cells -
PT particularly for controlling development of Th1 and Th2 cells for
PT treatment of allergy, autoimmune disease etc.
XX
XX Claim 31; Fig 11; 151pp; English.
XX
CC This is the amino acid sequence of the mouse NF-AT Interacting Protein 45
CC (NIP45). The gene sequence was isolated by using a yeast two-hybrid
CC detection system for proteins that interact with the NF-AT Rel homology
CC domain (RHD). The assay used, as a "bait", a 900 bp fragment of the
CC murine NF-ATP encoding the region spanning amino acids 228-250. NF-AT is
CC a multisubunit transcription complex containing a cyclosporin A sensitive
CC cytoplasmic phosphoprotein and an inducible component of the AP-1 family
CC of transcription factors. The screen was carried out on a cDNA library
CC prepared from the murine T cell line D10. One class of proteins, NIP45
CC designated NIP45, bound the NF-AT-RHD region with high affinity. NIP45
CC can be used in a claimed method to inhibit or stimulate production of
CC NF-AT family protein, particularly interleukin-4 (IL-4) and the treated
CC cells may be administered to control development of Th1 or Th2 cells by
CC modulating the production of a T-helper type 2 associated cytokine.
CC Especially the method is useful in the inhibition of Th2 in allergy,
CC cancer or infections, and promotion of Th2 in autoimmune disease and
CC transplantation.
CC
SQ Sequence 412 AA:

Query Match 100.0%; Score 2099; DB 18; Length 412;
Best Local Similarity 100.0%; Pred. No. 9.5e-175;
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEPLRGPRSRGCGARGARGARCGRPAROSPARIIPDTVLVLDVSDSEVLEVAD 60
DB 1 maepilrgprsrgrgarrargarrargrparqsparrilpdtvlvldvsdsdeevleavd 60
QY 61 PVEVPVAPRLPAPAKPEQSDSDSEGAEGPAGAPRTLVRRRRRLDPGEAPVVPYSGK 120
DB 61 pvevpvparlpapakpeqdsdsdsegaegpagaprtlvrrrrrllldpgeapvvpysgk 120
QY 121 VOSSLNLIPDNSSLKLCGSEPEDEADLTNCGSPSEDDALPGSGWRKKLRKCKEKEK 180
DB 121 vsslnl1pdnssl1klcspedeaddltncgsspseddalpsgswrkk1lkkckeek 180
QY 181 KMEEPDODISPLPOPSRNKSRKHTEALOKLREVNKRLODDRSCSPHOHSPALOSTND 240
DB 181 kmeefpdqdislpqpsrnrksrkhteaqlkrewnkrlqddirsc1spkqhspsalqstd 240
QY 241 DEVVLEGPVLPOSSRLFTLTKIRCADLVRLPYRMSEPLQNTVVDHMANHLGVSPNRILL 300
DB 241 devvlevegvlpqssrlftltkirradlvrlpyrmseplqnvvdhmanhlgvsnrilll 300
QY 301 FGSELSPTATPSTLKLGVADITDCVYLAASSSEATETSGELRLRQGGKHKHMLEISLSP 360
DB 301 fgeselsptatpstlklgyaditdcvylasssaatetsgellrlrygkxhqmleislsp 360
QY 361 DSPILKVLMSHYEAMGLSGHKLSFPFDGTRKLSGKELPADLGESGDLIEWWG 412
DB 361 dspilkvlmshyeamglsghklsffpdgtrklsgkelpadlglesgdliewwg 412

RESULT 2
AA08331
ID AAY08331 standard; protein; 412 AA.
XX
XX AAY08331;
AC
XX
XX 13-JUL-1999 (first entry)
XX
XX Mouse NIP45 protein.
```

```
XX NIP45; mouse; transcription activator; IL-4; interleukin-4;
KW NF-AT interacting protein-45; modulator; treatment; inflammation;
KW autoimmune disease; HTP screening; drug testing; allergic therapy;
KW T-cell dependent response; cytokine; diagnostic; immunosuppressant.
XX
XX Mus musculus.
XX
XX WO9921993-A1.
XX
XX 06-MAY-1999.
XX
XX 21-OCT-1998; 98WO-GB03141.
XX
XX 24-OCT-1997; 97GB-0022388.
XX
XX (ZENE ) ZENECA LTD.
XX
XX Liu D, Zhao J, Zhou H;
XX
XX WPI, 1999-312964/26.
XX
XX Polynucleotides encoding human NF-AT interacting protein (NIP45)
XX
XX Disclosure; Fig 4; 84pp; English.
XX
XX This invention describes a novel human NF-AT interacting protein-45
XX NIP-45. Human NIP45 is a transcriptional trans-activation factor of the
XX interleukin 4 (IL-4) gene. Modulators of human NIP45 and IL-4 can be used
XX to treat conditions mediated by NIP45 or IL-4, e.g. inflammation or
XX autoimmune disease. NIP45 is a good candidate target for HTP screening
XX and/or testing system for drugs that will alleviate T-cell dependent
XX autoimmune and allergic responses, and for cytokine-based therapies of
XX chronic disease. Expression of NIP45 can be inhibited, and IL-4
XX expression modulated, by administering an antisense molecule. Antibodies
XX against NIP45 and primers can be used in diagnostic assays. Discovery of
XX a human NIP45 is advantageous in that it provides the ability to control
XX IL-4 transcription, which is of importance for anti-inflammation and
XX immunosuppressant drug development.
XX
SQ Sequence 412 AA:

Query Match 100.0%; Score 2099; DB 20; Length 412;
Best Local Similarity 100.0%; Pred. No. 9.5e-175;
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEPLRGPRSRGCGARGARGARCGRPAROSPARIIPDTVLVLDVSDSEVLEVAD 60
DB 1 maepilrgprsrgrgarrargarrargrparqsparrilpdtvlvldvsdsdeevleavd 60
QY 61 PVEVPVAPRLPAPAKPEQSDSDSEGAEGPAGAPRTLVRRRRRLDPGEAPVVPYSGK 120
DB 61 pvevpvparlpapakpeqdsdsdsegaegpagaprtlvrrrrrllldpgeapvvpysgk 120
QY 121 VOSSLNLIPDNSSLKLCGSEPEDEADLTNCGSPSEDDALPGSGWRKKLRKCKEKEK 180
DB 121 vsslnl1pdnssl1klcspedeaddltncgsspseddalpsgswrkk1lkkckeek 180
QY 181 KMEEPDODISPLPOPSRNKSRKHTEALOKLREVNKRLODDRSCSPHOHSPALOSTND 240
DB 181 kmeefpdqdislpqpsrnrksrkhteaqlkrewnkrlqddirsc1spkqhspsalqstd 240
QY 241 DEVVLEGPVLPOSSRLFTLTKIRCADLVRLPYRMSEPLQNTVVDHMANHLGVSPNRILL 300
DB 241 devvlevegvlpqssrlftltkirradlvrlpyrmseplqnvvdhmanhlgvsnrilll 300
QY 301 FGSELSPTATPSTLKLGVADITDCVYLAASSSEATETSGELRLRQGGKHKHMLEISLSP 360
DB 301 fgeselsptatpstlklgyaditdcvylasssaatetsgellrlrygkxhqmleislsp 360
QY 361 DSPILKVLMSHYEAMGLSGHKLSFPFDGTRKLSGKELPADLGESGDLIEWWG 412
DB 361 dspilkvlmshyeamglsghklsffpdgtrklsgkelpadlglesgdliewwg 412
```


XX		03-SEP-2001	(first entry)	
DT				
XX				
DE		Human colon cancer antigen protein SEQ ID NO:5637.		
XX				
KW		Human; colon cancer; colon cancer antigen; diagnosis; detection;		
XX		colorectal carcinoma.		
OS	Homo sapiens.			
PN	MO200122920-A2.			
XX				
PD	05-APR-2001.			
PE	28-SEP-2000; 2000WO-US26524.			
PR	29-SEP-1999; 99US-0157137.			
PR	03-NOV-1999; 99US-0165280.			
PA	(HUMA-) HUMAN GENOME SCI INC.			
PI	Ruben SM, Barash SC, Birse CE, Rosen CA;			
DR	WPI: 2001-235357/24.			
XX	N-PSDB: AAH34478.			
PT	Nucleic acids encoding 4277 human colon cancer-associated polypeptides,			
XX	useful for preventing, diagnosing and/or treating colorectal cancers -			
PS	Claim 11; Page 7351-7353; 9803pp; English.			
XX				
CC	AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon			
CC	cancer-associated nucleic acid molecules (N) and proteins (P), where			
CC	the proteins are collectively known as colon cancer antigens. The colon			
CC	cancer antigens have cytosolic activity and can be used in gene			
CC	therapy and vaccine production. N and P may be used in the prevention,			
CC	diagnosis and treatment of diseases associated with inappropriate p			
CC	expression. For example, N and P may be used to treat disorders			
CC	associated with decreased expression by rectifying mutations or deletions			
CC	in a patient's genome that affect the activity of P by expressing			
CC	inactive proteins or to supplement the patients own production of P.			
CC	Additionally, N may be used to produce the colon cancer-associated Ps,			
CC	by inserting the nucleic acids into a host cell and culturing the cell			
CC	to express the proteins. N and P can be used in the prevention, diagnosis			
CC	and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204			
CC	and AAH7789 represent sequences used in the exemplification of the			
CC	present invention.			
CC	N.B. Pages 666 to 682 and page 7053 of the sequence listing were			
CC	missing at time of publication, meaning no sequences are present for			
CC	SEQ ID NO:1027 to 1052, 7921 and 7922.			
XX				
SQ	Sequence 555 AA:			
	Query Match 5.9%; Score 124.5; DB 22; Length 555;			
	Best Local Similarity 23.1%; Pred. No. 0.029;			
	Matches 93; Conservative 41; Mismatches 143; Indels 125; Gaps 20;			
OY	8 RPPRRGGGARRARCARGRCPRAROSPRLIPDYLVLY-----SDSD 52			
Db	: : : : :			
OY	97 rrprrrpglrygrrkkk---kapsasdsdkadsgakpavamarasssssssdpt 151			
OY	53 EEWLE-----VADVEVPVARLPPARP-----EDDSDD-----SEGAABEPAPAPT 96			
Db	: : : : : : : :			
OY	152 vavvkxkpriqrkpaekplphprgrkpkperippsssssdsvdrisekwrtdearrtel 211			
OY	lvvrrrr-----rrl-----ldpegavrvpyvsqkvosslnlripdnnslrk 136			
Db	: : : :			
OY	212 earrirregeeeelrrlrleagekeekererrradrgae-----rgsggsdelreddepvx 266			
OY	137 -----LCSSPED--EALUTNSGSSSPSEDALPGSSPMRKRLR----- 172			
Db	267 kgvgkgvggypgpssdsdpesaeleraaksakkpqsstaparkpqgkekrrvpeekqga 326			

Oy	173	--KICEEEKKMEEP-DODISPLFQPSRNKSKRHTEALQKLK-RVNNKFLQ---- <td>22</td>	22
Db	327	kpvkvevtrkrsegiismdtrkvekkkepsveek-----lqklhseikfalkvdsdvkr	37
Oy	225	CLS-----PKQIQSAPLSTFDEVLVE-----GPVLPOSSRLFTLKIRCRADLVR	27
Db	380	clnleelgltqstqlqknidvatlkikrirykankdmekaeayt-----r	42
Oy	271	LPRMSEPLQNVYDHMANHLGVSPNRILLGESELSPTATP	312
Db	430	lksrvllgpklaevqkv-nkagmekaeeklagelaegeep	470
RESULT 9			
AAI99426			
ID	AAI99426 standard; Protein; 671 AA.		
XX	AAI99426;		
AC			
XX			
DT	08-AUG-2000 (first entry)		
XX			
DE	Human PRO1604 (UNQ785) amino acid sequence SEQ ID NO:308.		
XX			
KW	Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;		
KM	transmembrane; secretion; immunoadhesion; pharmaceutical; screening.		
XX			
OS	Homo sapiens.		
XX			
PM	WO200012708-A2.		
XX			
PD	09-MAR-2000.		
XX			
PF	01-SEP-1999; 99WO-US20111.		
XX			
PR	01-SEP-1998;	98US-0098716	
PR	01-SEP-1998;	98US-0098749	
PR	01-SEP-1998;	98US-0098750	
PR	02-SEP-1998;	98US-0098803	
PR	02-SEP-1998;	98US-0098821	
PR	02-SEP-1998;	98US-0098843	
PR	09-SEP-1998;	98US-0099536	
PR	09-SEP-1998;	98US-0099596	
PR	09-SEP-1998;	98US-0099598	
PR	09-SEP-1998;	98US-0099602	
PR	09-SEP-1998;	98US-0099642	
PR	10-SEP-1998;	98US-0099741	
PR	10-SEP-1998;	98US-0099754	
PR	10-SEP-1998;	98US-0099763	
PR	10-SEP-1998;	98US-0099792	
PR	10-SEP-1998;	98US-0099808	
PR	10-SEP-1998;	98US-0099812	
PR	10-SEP-1998;	98US-0099815	
PR	10-SEP-1998;	98US-0099816	
PR	15-SEP-1998;	98US-0100385	
PR	15-SEP-1998;	98US-0100388	
PR	15-SEP-1998;	98US-0100390	
PR	16-SEP-1998;	98US-0100584	
PR	16-SEP-1998;	98US-0100627	
PR	16-SEP-1998;	98US-0100661	
PR	16-SEP-1998;	98US-0100662	
PR	16-SEP-1998;	98US-0100664	
PR	17-SEP-1998;	98US-0100683	
PR	17-SEP-1998;	98US-0100684	
PR	17-SEP-1998;	98US-0100710	
PR	17-SEP-1998;	98US-0100711	
PR	17-SEP-1998;	98US-0100919	
PR	17-SEP-1998;	98US-0100930	
PR	18-SEP-1998;	98US-0100848	
PR	18-SEP-1998;	98US-0100849	
PR	18-SEP-1998;	98US-0101014	
PR	18-SEP-1998;	98US-0101068	
PR	18-SEP-1998;	98US-0101071	


```
ID AAB66175 standard; protein; 671 AA.
XX
XX AAB66175;
XX
XX 02-APR-2001 (first entry)
XX
XX Protein of the invention #87.
XX
XX Secreted; transmembrane; gene therapy.
XX
XX Unidentified.
XX
XX WO200078961-A1.
XX
XX 28-DEC-2000.
XX
XX 18-FEB-2000; 2000WO-US04342.
XX
XX 23-JUN-1999; 99US-0141037.
XX
XX 20-JUL-1999; 99US-0144758.
XX
XX 26-JUL-1999; 99US-0145698.
XX
XX 01-SEP-1999; 99WO-US20111.
XX
XX 29-OCT-1999; 99US-0162506.
XX
XX 30-NOV-1999; 99WO-US28313.
XX
XX 02-DEC-1999; 99WO-US28531.
XX
XX 16-DEC-1999; 99WO-US30095.
XX
XX 05-JAN-2000; 2000WO-US00219.
XX
XX 06-JAN-2000; 2000WO-US00376.
XX
XX (GENTH ) GENENTECH INC.
XX
XX Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S,
XX PI Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gunney AL, Hillan KJ,
XX PI Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D,
XX PI Matanabe CK, Williams PM, Wood WI;
XX
XX WPI: 2001-071395/08.
XX
XX Secreted and transmembrane proteins and nucleic acids designated PRO,
XX PT useful as hybridization probes, in chromosome and gene mapping and gene
XX therapy -
XX
XX Claim 1; Fig 174; 787pp; English.
XX
XX The present invention relates to secreted and transmembrane proteins.
XX CC These proteins and the DNA encoding them may be used as hybridization
XX CC probes, in chromosome and gene mapping and in the generation of
XX CC anti-sense RNA and DNA. They may also be used to generate either
XX CC transgenic animals or knockout animals which are in turn useful for
XX CC development and screening of therapeutically useful reagents.
XX CC The nucleic acids may also be used in gene therapy.
XX
XX Sequence 671 AA:
SQ
Query Match 5.9%; Score 124.5; DB 22; Length 671;
Best Local Similarity 23.1%; Pred. No. 0.036;
Matches 93; Conservative 41; Mismatches 143; Indels 125; Gaps 20;
```

```
DB 363 krgtkgrrppssdseseaelereakksakbkpgssteparkpgkcvrpeekqga 442
QY 173 --KCKEKEKKMEFP-DODISPLPQSSRNKSRKHTALQKLR-EVKKRLQ---DLRS 224
DB 443 kpvkvertrkrsefgsmarkvekkkepsveek-----lqklhselkftalkvdspxdkr 495
QY 225 CLS-----PKOHSPALQSTDDEVLYE-----GPVLPOSSRLFTLIRCRADLVR 270
DB 496 clnaleejglqtvtsglqkntdvatlkkrtrykankdvmekeaeyt-----r 545
QY 271 LPRMSEPLQNVYDHMANHIGVSPNRILLPGSESLPTAMP 312
DB 546 lksrvlgpkieavqkv-nkagmekeakeeklagaeelageeap 586
RESULT 11
AAY33298
ID AAY33298 standard; Protein; 950 AA.
XX
XX AAY33298;
XX
XX 26-NOV-1999 (first entry)
XX
XX Human membrane spanning protein MSP-5.
XX
XX Membrane spanning protein; treatment; diagnosis; neoplastic disorder;
XX KM prevention; human; immunological disorder; reproductive disorder; MSP-5.
XX
XX Homo sapiens.
XX
XX WO9946380-A2.
XX
XX 16-SEP-1999.
XX
XX 09-MAR-1999; 99WO-US05073.
XX
XX 13-MAR-1998; 98US-0039064.
XX
XX (INCY-) INCYTE PHARM INC.
XX
XX Tang YT, Bandman O, Lai P, Hillman JL, Yue H, Corley NC;
XX PI Guegler KJ, Kaser MR, Baughn MR, Shah P;
XX
XX WPI: 1999-551409/46.
XX
XX DR N-PSDB; AAZ09839; AAZ09840.
XX
XX New human membrane spanning proteins used to, e.g. prevent and treat
XX PT neoplastic disorders -
XX
XX Claim 1; Page 73-76; 81pp; English.
XX
XX This invention describes novel human membrane spanning proteins (MSPs),
XX CC and the polynucleotides encoding them. The products of the invention are
XX CC used to diagnose, prevent and treat neoplastic, immunological and
XX CC reproductive disorders. This sequence represents the human membrane
XX CC spanning protein MSP-5.
XX
XX Sequence 950 AA:
SQ
Query Match 5.9%; Score 123; DB 20; Length 950;
Best Local Similarity 19.9%; Pred. No. 0.085;
Matches 108; Conservative 74; Mismatches 203; Indels 158; Gaps 23;
```



```

Db 433 pskatsiesftnapptlisepstraagpgrfrdmrdreghndpvnv---gppkpak 489
QY 126 NLIPDNSLKLCPSEPEDADLTNSSSPEDDALPSGSPMKR-----KLRKCEK 177
Db 490 ekppkkaqdkllsneyekydistrasqledelgvgnpklkakeskhekpeke 549
QY 178 EEKME-EFPDODISPLPQPSRRNRKSRHTEALQKLEVNKRLOD-----221
Db 550 kkkkmknendkllksekgnkkeskqekskkkgkktgedgyqktnkhtfspx 609
QY 222 -----LRSCISPKOHOSPALQSTDD-----EVLVEGPVLPQ 253
Db 610 ksavadllgsfegkrillllltapkaennmyvgqdeylestfckmatrkisvltfgyv---666
QY 254 SSRLFTKICGRADIVRL-----PYRM--SEPL--QNVYDHMANLIGVSPRILLPFEESE 305
Db 667 --nsmtnk1---dhfqldaekpmrvvddedlvdqrliselrkeygntlyndftmvltdvd 720
QY 306 LSPY---ATPSTKLGVADIIDCVVLASSSEATEETSOELRLRVQGE-KHQMLEISLSP- 360
Db 721 lrvkgyeyvpltnk-svfdltd-----tfgsrikdmekqkgeivckekdkqslentfsrf 775
QY 361 -----DSPKLVMKSHYEAMGLSGHKLSP---FFDGTKLSGKELPADLGLSGDLI 408
Db 776 rwrtrllivisaapedwayeqslalsgqacnfglrhltilklilg-----vgeevgavl 829
QY 409 EVW 411
Db 830 ell 832

RESULT 12
AAM53863
ID AAM53863 standard; peptide; 1780 AA.
XX
AC AAM53863;
XX
DT 13-JUL-1998 (first entry)
XX
DE Human gravin polypeptide.
XX
KW Gravin; kinase anchoring protein; type II regulatory subunit; PKA; PKC;
KM CAMP-dependent protein kinase; protein kinase C; autoimmune disease;
KM Myasthenia gravis; nicotinic acetylcholine receptor.
XX
OS Homo sapiens.
XX
PN US741890-A.
XX
PD 21-APR-1998.
XX
PF 19-DEC-1996; 96US-0769309.
XX
PR 19-DEC-1996; 96US-0769309.
XX
PA (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI Klaufack TM, Nauert JB, Scott JD;
XX
DR WPI; 1998-260552/23.
XX
DR N-PSDB; AAV23545.
XX
PT New polypeptide fragments of protein kinase binding protein gravin -
PT are useful for the study of modulation of action between gravin and
PT protein kinase(s)
XX
PS Example 1; Column 19-32; 32pp; English.
XX
CC This sequence corresponds to the human gravin polypeptide, and represents
CC a polypeptide of the invention. The polypeptides are fragments capable of
CC binding to type II regulatory subunit of CAMP-dependent protein kinase
CC (PKA). Gravin is a kinase anchoring protein that binds to type II

```

```

CC regulatory subunits of PKA and protein kinase C (PKC). Gravin is also an
CC antigen of the autoimmune disease Myasthenia gravis (MG), where a patient
CC develops antibodies against their own nicotinic acetylcholine receptors.
CC The polypeptides are useful for providing analogues of gravin in the
CC study of the modulation (e.g. blocking, inhibiting and stimulating) of
CC interactions between gravin and kinase. The peptides are involved in the
CC modulation of gravin-kinase interactions.
XX
SQ Sequence 1780 AA.

Query Match 5.88; Score 122.5; DB 19; Length 1780;
Best Local Similarity 22.08; Pred. No. 0.24;
Matches 93; Conservative 62; Mismatches 152; Indels 115; Gaps 22;

QY 40 PDYV-LYDVSDSEDEVLVADPVEVYARLPAPAK---PEQDSDSDEGAEPACAPR 95
Db 192 pdtvglltvkkdegeagaagdhgdpislgaeascksepkxstekpeee-----240
QY 96 TLVRRRRRLIDPGEAPVVPYSGKVQSLNLIPONSLKLCPSEPED--EADLTNSGS 153
Db 241 tlkregshaelsp-----paesgqa-----veeckeegeekqekesksae 281
QY 154 SPSEDDALPSGSP-----WRKK--LRKCKEKE---EKKMEFPDODISPLPQP 196
Db 282 sptspvtsetgsttkkfttgwgagwrktsfrkpkedeveaseekkeqepk-----vdt 336
QY 197 SSRKRSRKHTEALOKLAEVNKRLODLRSCLSPK---QHOSALOSTDDEVVLVEGP----249
Db 337 eedqkaevaseklaseaqahpq-epaesahprlisaeyekvlpse-eeqysgsgpseek 394
QY 250 VLPOSSRLFTKICGRADIVRLPYRMSEPLQNVYDHMANHLGVSPNRILLFGSELSPT 309
Db 395 paplatevfdckievheevaeavhsvteerteeg-----kteveet 437
QY 310 A--TPSTKLGVADIIDCVVLASSSEATEETSOELRLR---VQGEKQMLEISLSPDPL 364
Db 438 agsvpaeealvgmd-----aepqaeapakelyklketcvsgedpfgad--lspdekx 487
QY 365 -----KVLMSHYEAMGLSGHKLSPFPDGT---KLSGKELPADLG--LESGLDI 408
Db 488 lskppegvsevevlssqetmkgvgsplkllftstglklisqkxqkkrvggdeesgeht 547
QY 409 EV 410
Db 548 qv 549

RESULT 13
AAB15380
ID AAB15380 standard; Protein; 1780 AA.
XX
AC AAB15380;
XX
DT 26-JAN-2001 (first entry)
XX
DE Human gravin protein sequence.
XX
KW Human; gravin; PKA RII binding site; myasthenia gravis;
KM kinase anchoring protein; CAMP dependent protein kinase.
XX
OS Homo sapiens.
XX
FT Key Location/Qualifiers
FT Binding-site 265..556
FT Binding-site /note="PKC binding site"
FT Binding-site 1526..1582
FT /note="PKA RII binding site"
FT Region 1537..1563
XX
XX /note="PKA anchoring site"
XX
PN US6090929-A.

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 22, 2002, 10:06:15 ; Search time 19.24 Seconds
(without alignments)
481.879 Million cell updates/sec

Title: US-09-617-923-2

Perfect score: 2099

Sequence: 1 MAEPLRGGRGPRGRCGARR.....GKELPADLGLESGDLIEWWG 412

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/ECTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfilest1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2099	100.0	412	2	US-08-755-584-2
2	2099	100.0	412	3	US-09-192-611-2
3	122.5	5.8	1780	1	US-08-769-309A-5
4	122.5	5.8	1780	3	US-08-994-570-5
5	120.5	5.7	565	4	US-08-961-083-218
6	113.5	5.4	661	2	US-08-795-868-14
7	112.5	5.4	95	4	US-09-100-802-10
8	111	5.3	897	1	US-08-095-737-4
9	111	5.3	897	2	US-08-480-145-4
10	111	5.3	897	2	US-08-477-389-4
11	109.5	5.2	1142	4	US-09-061-709-2
12	108.5	5.2	504	4	US-08-955-918C-7
13	108.5	5.2	505	1	US-08-631-200-2
14	108.5	5.2	505	1	US-08-630-592-4
15	108.5	5.2	505	1	US-08-714-991-4
16	108.5	5.2	505	2	US-08-829-553-2
17	108.5	5.2	505	2	US-08-922-267A-2
18	108.5	5.2	505	2	US-08-936-707A-2
19	108.5	5.2	505	2	US-08-936-706A-2
20	108.5	5.2	505	3	US-09-248-203-2
21	108.5	5.2	505	3	US-09-032-365A-4
22	108.5	5.2	505	4	US-08-812-824-3
23	108.5	5.2	505	4	US-09-406-071-2
24	108	5.1	459	1	US-08-630-592-2
25	108	5.1	459	1	US-08-714-991-2
26	108	5.1	459	3	US-09-032-365A-2
27	108	5.1	673	2	US-08-435-073A-6

28	108	5.1	1001	4	US-09-060-410-2	Sequence 2, App11
29	107	5.1	521	2	US-08-721-684C-2	Sequence 2, App11
30	107	5.1	521	2	US-09-005-970-2	Sequence 2, App11
31	107	5.1	521	4	US-09-407-715-2	Sequence 2, App11
32	107	5.1	570	3	US-08-826-246-2	Sequence 2, App11
33	107	5.1	570	3	US-08-944-495-2	Sequence 2, App11
34	107	5.1	570	3	US-09-126-640-7	Sequence 2, App11
35	107	5.1	570	4	US-08-925-588-2	Sequence 2, App11
36	106.5	5.1	723	2	US-08-548-159-5	Sequence 5, App11
37	106.5	5.1	1142	2	US-08-993-118-7	Sequence 7, App11
38	106.5	5.1	1142	3	US-08-845-528C-7	Sequence 7, App11
39	104.5	5.0	861	1	US-08-484-105-18	Sequence 18, App1
40	104.5	5.0	861	1	US-08-484-106-18	Sequence 18, App1
41	104	5.0	126	2	US-08-853-974-1	Sequence 1, App11
42	104	5.0	126	4	US-09-172-988-1	Sequence 1, App11
43	104	5.0	3248	1	US-08-353-700-1	Sequence 1, App11
44	104	5.0	3248	5	PCT-US95-16216-1	Sequence 1, App11
45	103.5	4.9	586	2	US-08-630-822A-70	Sequence 70, App1

ALIGNMENTS

RESULT 1
US-08-755-584-2
; Sequence 2, Application US/08755584
; Patent No. 5858711
GENERAL INFORMATION:
; APPLICANT: Gilmcher, Laurie H.
; APPLICANT: Hodge, Martin R.
; TITLE OF INVENTION: NF-AT-INTERACTING PROTEIN NIP4S AND METHODS
; NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/755,584
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Giulio A. Decontini, Jr.
REGISTRATION NUMBER: 31,503
REFERENCE/DOCKET NUMBER: HUI-026
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 412 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-755-584-2

Query Match 100.0%; Score 2099; DB 2; Length 412;
Best Local Similarity 100.0%; Pred. No. 1.8e-174;
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAEPLRGGRGPRGRCGARRGRCPRAROSPRLIPDTVLVDLVSDSEVLEVAD 60

```
|||||
Db 1 MAEPLRGKPRSGRGARARGARCPRARQSPARLLPDTVLVDVSDSEVLEVAD 60
QY 61 PVEVPARLPAPAKPEODSDSDEGAAGPAGAPRTLVRRRRRLDPGEAVPVVYSGK 120
Db 61 PVEVPARLPAPAKPEODSDSDEGAAGPAGAPRTLVRRRRRLDPGEAVPVVYSGK 120
QY 121 VOSSLNLIPDNSSLKLCPCSEPEDEADLTNCGSSPSEDDALPGSPWRRKLRKCKEERK 180
Db 121 VOSSLNLIPDNSSLKLCPCSEPEDEADLTNCGSSPSEDDALPGSPWRRKLRKCKEERK 180
QY 181 KMEEPDODISPLPQSSRNKSRKHTKTEALQKREVNKRLODLRSLCSFKOHSPALQSTD 240
Db 181 KMEEPDODISPLPQSSRNKSRKHTKTEALQKREVNKRLODLRSLCSFKOHSPALQSTD 240
QY 241 DEVVLEGEVLPQSSRLFTLTKIRCRADLVRLPVRMSEPLQNVVDHMANHLGVSPPNRILL 300
Db 241 DEVVLEGEVLPQSSRLFTLTKIRCRADLVRLPVRMSEPLQNVVDHMANHLGVSPPNRILL 300
QY 301 FGESELSPTATPSTLTGLGADIIDCVYLAASSSEATETSOELRLRVQGEKHKOMLEISLSP 360
Db 301 FGESELSPTATPSTLTGLGADIIDCVYLAASSSEATETSOELRLRVQGEKHKOMLEISLSP 360
QY 361 DSPLKVLMSHYEAMGLSGHKLSFFPDGTKLSGKELPADLGLESDDLIEWMG 412
Db 361 DSPLKVLMSHYEAMGLSGHKLSFFPDGTKLSGKELPADLGLESDDLIEWMG 412

RESULT 2
US-09-192-611-2
; Sequence 2, Application US/09192611
; Patent No. 6090561
; GENERAL INFORMATION:
; APPLICANT: Glimcher, Laurie H.
; APPLICANT: Hodge, Martin R.
; TITLE OF INVENTION: NF-AT-INTERACTING PROTEIN NIP45 AND METHODS
; TITLE OF INVENTION: OF USE THEREFOR
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/192,611
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/755,584
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Giulio A. Decontal, Jr.
; REGISTRATION NUMBER: 31,503
; REFERENCE/DOCKET NUMBER: HUI-026
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 412 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-192-611-2
```

```
Query Match 100.0%; Score 2099; DB 3; Length 412;
Best Local Similarity 100.0%; Pred. No. 1,8e-174;
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEPLRGKPRSGRGARARGARCPRARQSPARLLPDTVLVDVSDSEVLEVAD 60
Db 1 MAEPLRGKPRSGRGARARGARCPRARQSPARLLPDTVLVDVSDSEVLEVAD 60
QY 61 PVEVPARLPAPAKPEODSDSDEGAAGPAGAPRTLVRRRRRLDPGEAVPVVYSGK 120
Db 61 PVEVPARLPAPAKPEODSDSDEGAAGPAGAPRTLVRRRRRLDPGEAVPVVYSGK 120
QY 121 VOSSLNLIPDNSSLKLCPCSEPEDEADLTNCGSSPSEDDALPGSPWRRKLRKCKEERK 180
Db 121 VOSSLNLIPDNSSLKLCPCSEPEDEADLTNCGSSPSEDDALPGSPWRRKLRKCKEERK 180
QY 181 KMEEPDODISPLPQSSRNKSRKHTKTEALQKREVNKRLODLRSLCSFKOHSPALQSTD 240
Db 181 KMEEPDODISPLPQSSRNKSRKHTKTEALQKREVNKRLODLRSLCSFKOHSPALQSTD 240
QY 241 DEVVLEGEVLPQSSRLFTLTKIRCRADLVRLPVRMSEPLQNVVDHMANHLGVSPPNRILL 300
Db 241 DEVVLEGEVLPQSSRLFTLTKIRCRADLVRLPVRMSEPLQNVVDHMANHLGVSPPNRILL 300
QY 301 FGESELSPTATPSTLTGLGADIIDCVYLAASSSEATETSOELRLRVQGEKHKOMLEISLSP 360
Db 301 FGESELSPTATPSTLTGLGADIIDCVYLAASSSEATETSOELRLRVQGEKHKOMLEISLSP 360
QY 361 DSPLKVLMSHYEAMGLSGHKLSFFPDGTKLSGKELPADLGLESDDLIEWMG 412
Db 361 DSPLKVLMSHYEAMGLSGHKLSFFPDGTKLSGKELPADLGLESDDLIEWMG 412

RESULT 3
US-08-769-309A-5
; Sequence 5, Application US/08769309A
; Patent No. 5741890
; GENERAL INFORMATION:
; APPLICANT: Scott, John D.,
; APPLICANT: Naurek, Brian J.,
; APPLICANT: Klauck, Theresa M.
; TITLE OF INVENTION: Protein Binding Domains of Gravin
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower/233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/769,309A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5741890and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/33451
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1780 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
```

US-08-769-309A-5

Query Match	5.8%;	Score 122.5;	DB 1;	Length 1780;
Best Local Similarity	22.0%;	Pred. No. 0.068;		
Matches 93;	Conservative 62;	Mismatches 152;	Indels 115;	Gaps 22

QY	40	PDTY-LVDLVSDSEEVYLEVADPVEVAVARLPAPAK--PEODSDSDSEGAEGPAGAPR	95
Db	192	PDTYQLLVTKKDEBEGAAGDHDPPSLGGEAASKSESPQSTENKEDEE-----	240
QY	96	TLVYRRRRRLIDPCEAPVPVYSGKVOSSLNLIPDNSSLKLCPSPEPD--EADLTNNGS	15
Db	241	TLKEGSHAEISp-----PAESGOA-----VEECKEEBEEKOEKESPSCAE	281
QY	154	SPSDDALPSCSP-----WRKK--LRKCKEKE--EKKMEFPPODLSPLPQ	196
Db	282	SPSPVYSENGSTTKKFFFTQGMAGMRKRTSPRKPEDEVEASEKKKQDEPK-----VDT	333
QY	197	SSRNKSRKHTEALOKLREVNKRLODLKSLSPK---OHOSFALOSTDEVLVEGP---	249
Db	337	EEDGKAFAVASGLFASDAQHQ--EPASAHPRSLATAYEKELPS--DEQVSGSGCPSEBK	399
QY	250	VLPOSSRLFTLKICRCADLYLRPYRMSEPLQNVVDHMANHLGVSNNRILLFGESELSP	309
Db	395	PAPLATEFEFDEKIEVHOEEVAEVAHVSTVEERTEO-----KTEVEET	437
QY	310	A--TPSTLKIGVADIIIDCVYLASSSEKFTSPSOELRL---VOGKERHOMLEISPSPL	366
Db	438	AGSVPAEELVGM-----AEQOEAPAKELVKLETCVSGGDPPQAGD--LSPBEKV	487
QY	365	-----KVLMSHYEAMGLSGHLSFFEDT---KLSGKELPADLG---LESODLI	408
Db	488	LSRPPEGVSEVLEKLSQGERKAYGSPCLKLFTSTGLKTLKSGKKQKKRGGDSESEHT	547
QY	409	EY 410	
Db	548	QY 549	

RESULT 4
 US-08-994-570-5
 Sequence 5, Application US/08994570
 Patent No. 6090929
 GENERAL INFORMATION:
 APPLICANT: Scott, John D.,
 APPLICANT: Nauret, Brian J.,
 APPLICANT: Klauack, Theresa M.
 TITLE OF INVENTION: Protein Binding Domains of Gravin
 NUMBER OF SEQUENCES: 24
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: Marshall, O'Toole, Garstein, Murray & Borun
 STREET: 6300 Sears Tower/233 South Wacker Drive
 CITY: Chicago
 STATE: Illinois
 COUNTRY: United States of America
 ZIP: 60606-6402
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/994,570
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: No. 6090929and, Greta E.
 REGISTRATION NUMBER: 35,302
 REFERENCE/DOCKET NUMBER: 27866/33451
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 312-474-6300
 TELEFAX: 312-474-0448

```

?       TELEX: 25-3856
?       INFORMATION FOR SEQ ID NO: 5
?       SEQUENCE CHARACTERISTICS:
?           LENGTH: 1780 amino acids
?           TYPE: amino acid
?           TOPOLOGY: linear
?       MOLECULE TYPE: protein
?       US-08-994-570-5

```

Query Match 5.8%; Score 122.5; DB 3; Length 1780;
Best Local Similarity 22.0%; Pred. No. 0.068;
Matches 93; Conservative 62; Mismatches 152; Indels 115; Gaps 22;

QY	40	PDIV-LVDIVSDSEEVLEVADPVEVAVARLPAPAK---	PEODSDSEGAAGAPAGR	95	
Db	192	PDIVOLLTVKKDEBEGAAGDHPDPLSGEAAKSEPEKQSTERE	----	240	
QY	96	TLVRRRRRLIDPEAVPVPYSGKVOSSLNLIPDNSSLKLCPEPD-	EADLINGS	153	
Db	241	TLKEOSHAEISp-----PAESGA-----	VEECKEKEBEKEKEPSKAE	281	
QY	154	SPSEDDALPSCSP-----WRKK--LRKKCKE--	EKKMEFPDDDISLPQ	166	
Db	282	SPSPVMSSETGSTRKFFTOGMAWRKKTSPRKKEDVEASEKKQDEPEK	-----VDT	336	
QY	197	SSRKSRRKHTALOKLEVRNRLQDLSCLSPK---OHQSPALQSTDEVVLVEGP-	249		
Db	337	EEDGKAEVASPKLNASQAHQ--EPASAHBPRLSALEYKVELPS--	BEQVSGSGPSEK	394	
QY	250	VLPOSSRLFTIKICRADVLRPLPRMSEPLQNVVDHNAHLGVSPNRILLFGESELSP	309		
Db	395	PAPLATEVEFEKEIVHOEEVAEVHVSVERTEEO-----	KTEVEET	437	
QY	310	A--PSTILCLGVADIICVVLASSSEAPTESOELRL--	VQGEKHOMELSLSPSL	364	
Db	438	AGSVPAEBELVGM-----AEQOEAPAKELKYLKETVSGEDPLQGD--	LSPEKY	487	
QY	365	-----KVLMSHYEAMGLSGHLSLFFEDT--	KLSGELPADLG--	LESGDLI	408
Db	488	LSKPPEGVSEVEMLSQGERMKVQGSPLKLTSTGLKLTSGKKOKKRGCGDESGENT	547		
QY	409	EY	410		
Db	548	QV	549		

```

RESULT      5
US-08-961-083-218
; Sequence 218, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:

```

FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 218:
SEQUENCE CHARACTERISTICS:
LENGTH: 565 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-218

Query Match 5.7%; Score 120.5; DB 4; Length 565;
Best Local Similarity 20.9%; Pred. No. 0.02;
Matches 81; Conservative 66; Mismatches 156; Indels 85; Gaps 19;

QY 29 PRARSP-----ARLIPDVLDVSDSEVLEVAAPVEV--PYARLP--APAKPE-Q 77
DB 34 PTEESPKKEPKSEVAKPTDDTLPKVEEGKEDSAPAEVPEEGEVESKPEEKVAAKPEEQ 93
QY 78 DSDSDEGAEGPAGAPRTLVRRRRRLDPGEAPVVPYSGVQSS--LNLIPDNSSL 134
DB 94 PDKKAESEKVGQAGEP-VAPREDEKAPVEPKQPEAPEEKAAVEETPKQESTPTDKAE 152
QY 135 LKLCPE-----SEPEDADLTNGS--SPSEDALPGSGFWRK 170
DB 153 EVFEPKEETVNOSEIQPKVETPAVEKQTEPEPKVEQAGEVAPREDQAPT-AVEPE 211
QY 171 LKKCKEKEKEEPPDDDISPLPPSSRNKSKHTEALQKREYVKRLQDRLRSCSPKQ 230
DB 212 KQPEVPEEKAAVE-----TPKPEDKIKGIGTEPEYDK-SELNNQI-DKASSSPID 261
QY 231 HOSPALQSTDEVEVLVEGVLPOSSRLFTLKIRCADLVRLPVRMSE--PLQNVYD--- 284
DB 262 YSTASYNAL-----GVLETAQGVYA-----SEPKQEPVNSETKLTALDALNV 307
QY 285 ---HANHLGVSPNRLLLFGESELSPTATPSTLKIGVADIIDCVLASSSEPTETSQSL 341
DB 308 DKTELNNITADKTKYKEHYSDRSWONIQTEVYTKAEKVAANTD---AQOSEVNEAVEKL 363
QY 342 RLVOGKEKHOMLEISLSPDSPKLKVLMS 369
DB 364 TATIE-----KLVELS---EKPLITLTS 383

RESULT 6
US-08-795-868-14
Sequence 14, Application US/08795868
Patent No. 5846773
GENERAL INFORMATION:
APPLICANT: Lee, Mu-En
APPLICANT: Hsieh, Chung-Ming
TITLE OF INVENTION: A SINGLE GENE ENCODING AORTIC-SPECIFIC
TITLE OF INVENTION: AND STRAINED-SPECIFIC MUSCLE CELL ISOFORMS AND USES THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/795,868
FILING DATE: 06-FEB-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/494,577
FILING DATE: 22-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 05433/032001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-342-5070
TELEFAX: 617-542-8906
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 661 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-795-868-14

Query Match 5.4%; Score 113.5; DB 2; Length 661;
Best Local Similarity 20.0%; Pred. No. 0.1;
Matches 103; Conservative 55; Mismatches 131; Indels 227; Gaps 25;

QY 4 PLRGR---GPRRG---KGRARRAGAR----- 25
DB 26 PARGHVHAPPSHSGACAARGHNRREAREYQVVALPGHRESRPOTPLSEASGRIMA 85
QY 26 -GRCPRAROSPARRLPDVTLVLDVSDSE--EVLEADPVEVYV-----ARLPAPAK 74
DB 86 LGRSRLVRAAGSRIL-----DKLOFEEERRSLRSDSPAPRLPWPVPLKRAARLEQPK 139
QY 75 PEQSDSDSEGAEGPAGAPRTLVRRR-----RRLDPGEAPVVPYS-----GK 120
DB 140 SERGAPWGTGPGASOELRAPGSAVERRRLLFOKAASLDERTQORSASDLELRFQOELGR 199
QY 121 VOSSLN---LIPDNSSL--LKLCPSEPE-DEADLTNNGSSPSSEDDAL----- 161
DB 200 IRRSTSEELVSHESLRTAQAPSPREPGEPLFSRSTPKTSRAASPAAPPPSS 259
QY 162 ---PSGSPWKRKLKRC-----EKEKKMEFP-----DQDISPLPQPS----- 197
DB 260 AKPQDEPGRRSRGPAGTTEPGEQDEVRRRDQPLRLRSRAIOECSPVPPPAADPE 319
QY 198 SRNKS---RKHTDALQKL-----EVNKRRLQ----- 220
DB 320 ARTKAPGRRKREPAAQAVFLPMATPGLGAAVPTQLEKNRAGPBAERLRKRGPEDEGPW 379
QY 221 ---DLRSLSPKQHO-----SPALQSTDEVYL-----VEGVLLQSSSLFTLKIRCAD 267
DB 380 GPWDRRGARSGCKGRAPRTSELSSDSYSAEBEPLADV----- 422
QY 268 LVRLPVRMSEPLQNVYVDHANHLGVSPNRLLLFGESELSPTATPSTLKIGVADIIDCV 327
DB 423 -----FELPLQNVY-----VAP-----GADVLLKCI 444
QY 328 LAS-----SSEATETSQELRLRVQCKEKHQML 354
DB 445 TANPPQVSWHKDGSALRSEGRLLRLRAEG-ERHTLL 479

RESULT 7
US-09-100-802-10
Sequence 10, Application US/09100802A
Patent No. 6294363
GENERAL INFORMATION:
APPLICANT: Madura, Kiran
TITLE OF INVENTION: Methods and Compositions for the Rapid

;; TITLE OF INVENTION: Purification of Proteasomes and Methods of Use of Components
;; FILE REFERENCE: UMDNJ97-11
;; CURRENT APPLICATION NUMBER: US/09/100,802A
;; EARLIER FILING DATE: 1998-06-19
;; NUMBER OF SEQ ID NOS: 17
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 10
;; LENGTH: 95
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Synthetic Sequence
US-09-100-802-10

Query Match 5.4%; Score 112.5; DB 4; Length 95;
Best Local Similarity 32.5%; Pred. No. 0.0085;
Matches 25; Conservative 18; Mismatches 33; Indels 1; Gaps 1;

OY 335 TETSQLRVRGCKEKNHLEISLSPDKVLMHYEAMGLSGHLSFFDGTLSGK 394
DB 12 TENNDINIKVAGD-GSYVOFKIKRHTPLSKLMKAYCERQISMRQIRFRFDGPINET 70
OY 395 ELPADLGESGLIEYW 411
DB 71 DTPAOLEMEDEDTIVF 87

RESULT 8
US-08-095-737-4
;; Sequence 4, Application US/08095737
;; Patent No. 5487979
;; GENERAL INFORMATION:
;; APPLICANT: DiFiore, Pier P
;; APPLICANT: Fazioli, Francesca
;; TITLE OF INVENTION: A Substrate for the Epidermal Growth
;; TITLE OF INVENTION: Factor Receptor Kinase
;; NUMBER OF SEQUENCES: 4
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Knobbe, Martens, Olson & Bear
;; STREET: 620 Newport Center Drive, Sixteenth Floor
;; CITY: Newport Beach
;; STATE: California
;; COUNTRY: United States of America
;; ZIP: 92660
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: IBM PC compatible
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/095,737
;; FILING DATE: 19930722
;; CLASSIFICATION: 530
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Israelsen, Ned A
;; REGISTRATION NUMBER: 29,655
;; REFERENCE/DOCKET NUMBER: NIH060,001A
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (619) 235-8550
;; TELEFAX: (619) 235-0176
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 897 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-095-737-4

Query Match 5.3%; Score 111; DB 1; Length 897;
Best Local Similarity 21.2%; Pred. No. 0.26;
Matches 92; Conservative 79; Mismatches 153; Indels 110; Gaps 24;

OY 61 PVEVPVAPRLP-----APA-KPEOD-----SDSDEGAAG-----PAGAPRTL 97
DB 202 PMSLPALVPVPSKRTVWVSPAERAKYDEIFLTKDKMDGYGSLVEYRTEFLTKGPSAL 261
OY 98 VRRRRRLDPGAPVVPVVS-----GKVQ-----SINLIPDNSSLK-----LC 138
DB 262 LAH-----INSLCDTRKCGKLSKDQFALAPHLI--NOKLTKGIDPPHSLT 304
OY 139 PS--EPDEADLTN--SGSSPSED-----DALPSGSPRRKRLRKCKEKKMEFPD 187
DB 305 PEMIPSDRSSLOKNITGSSPVADFSKIKELDTLNNETIVLQEKNNVEDLKEKEDTVK 364
OY 188 QDISPLPQSSRNKSRKHTALQKLRVNRLODLRSCLSPKQHQ-----SPALQSTDE 242
DB 365 QRTSEV--QDLQDEVOGESINLQKQKQOQVOELLGELDEQKAQLEQQLQEVRRKCAEE 422
OY 243 VLVY---EGVVLQSSKRLFTLK---ITCRADLVRLPYRMSEPIQNV-----VDHANHL 290
DB 423 AOLISSLKAEITQESQOISSYEELKARBELSLQOETQLEESVSGKAOLEPILOHL 482
OY 291 GVSPPNRLILFGSELSPTAT-----PSTLKLGVADITDCVVLASSSEATETSQ 339
DB 483 QESQOETSSQMOMLEKMDLETDNQSNWSSSPQSVLYNGATDY--CSLTSISSETANFNE 540
OY 340 --ELRLRVGCKEKNHLEISLSPD-SPLKVLMSHYEAMGLSGH-KLSFFDGTLSGKE 395
DB 541 HAQGNMLESEPTQOESSVNSPEIAPSDV--TDESAVTVAGNEKVTYPRFDDKHSKEE 598
OY 396 LPADLGESGLIE 409
DB 599 DP--FNVESSLTD 610

RESULT 9
US-08-480-145-4
;; Sequence 4, Application US/08480145
;; Patent No. 5717067
;; GENERAL INFORMATION:
;; APPLICANT: DiFiore, Pier P
;; APPLICANT: Fazioli, Francesca
;; TITLE OF INVENTION: A Substrate for the Epidermal Growth
;; TITLE OF INVENTION: Factor Receptor Kinase
;; NUMBER OF SEQUENCES: 4
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Knobbe, Martens, Olson & Bear
;; STREET: 620 Newport Center Drive, Sixteenth Floor
;; CITY: Newport Beach
;; STATE: California
;; COUNTRY: United States of America
;; ZIP: 92660
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: IBM PC compatible
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/480,145
;; FILING DATE: 07-JUN-1995
;; CLASSIFICATION: 530
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/095,737
;; FILING DATE: 22-JUL-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Israelsen, Ned A
;; REGISTRATION NUMBER: 29,655
;; REFERENCE/DOCKET NUMBER: NIH060,001A
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (619) 235-8550
;; TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 897 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-480-145-4

Query Match 5.3%; Score 111; DB 1; Length 897;
Best Local Similarity 21.2%; Pred. No. 0.26;
Matches 92; Conservative 79; Mismatches 153; Indels 110; Gaps 24;

QY 61 PVEVPVARLP-----APA-KPEQD-----SDSDEGAAG-----PAGAPRTL 97
DB 202 PMSLPALVPSPKRTWVSPAKAKYDEIFLTKDKMDGYSGLEVRETFKLTGLPSAL 261
QY 98 VRRRRRLDPGEAVPVVYS-----GKVO-----SLNLPDSSSLK-----LC 138
DB 262 LAH-----IWSLCDTRKGCCKLSKQDFALAFHLI--NQKLKIDPPHSLT 304
QY 139 PS--EPEDADLTN--SGSSPED-----DALPGSPWRKKLRKKCKEKKKEEPPD 187
DB 305 PEMIPSPDRSSLOKNITGSSPVADPSAIKELDTLNNEIYDLOREKNVNEODLKEKEDTVK 364
QY 188 QDISPLPQPSNRKSRKHTKLEVNKRLODRLSCISPKQH-----SPALQSTDE 242
DB 365 QRTSEV--QDLQDEVQRESINLQKLAQKQVQELLGELDEKQALDEQLQEVRRKCAEE 422
QY 243 VLVV--EGPVLPQSSRLFTLK---IRCRADVLRLPVRMSEPLQNV-----VDHMANHL 290
DB 423 AOLISLKAELITSOEISYEEELKAREELSRLOQETALESVESGKALQLEPQOHL 482
QY 291 GVSPPRIILLFGESELSPTAT-----PSTLKGVDIIDCVLASSSEATETSO 339
DB 483 QESQOETISSMQRLEKDKDETNNOSNMSSSPQSVLVNGATDY--CSLSTSSSETANFNE 540
QY 340 --ELRLVQKKEKHOMLEISLSPD--SPLKVLMSHYEAMGLSGH--KLSEFPDGTLSGKE 395
DB 541 HAEQGNLSEPTHQESSVRSSPEIAPSDV--TDESEAVTAGNEKVTYPRFDDKHSKEE 598
QY 396 LPADLGLESGLIE 409
DB 599 DP--FNVESSSLTD 610

RESULT 10
US-08-477-389-4
Sequence 4, Application US/08477389
Patent No. 5872219
GENERAL INFORMATION:
APPLICANT: DiFiore, Pier P
APPLICANT: Fazio, Francesca
TITLE OF INVENTION: A Substrate for the Epidermal Growth
TITLE OF INVENTION: Factor Receptor Kinase
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear
STREET: 620 Newport Center Drive, Sixteenth Floor
CITY: Newport Beach
STATE: California
COUNTRY: United States of America
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,389
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/095,737
FILING DATE: 22-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned A
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH060,001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 897 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-477-389-4

Query Match 5.3%; Score 111; DB 2; Length 897;
Best Local Similarity 21.2%; Pred. No. 0.26;
Matches 92; Conservative 79; Mismatches 153; Indels 110; Gaps 24;

QY 61 PVEVPVARLP-----APA-KPEQD-----SDSDEGAAG-----PAGAPRTL 97
DB 202 PMSLPALVPSPKRTWVSPAKAKYDEIFLTKDKMDGYSGLEVRETFKLTGLPSAL 261
QY 98 VRRRRRLDPGEAVPVVYS-----GKVO-----SLNLPDSSSLK-----LC 138
DB 262 LAH-----IWSLCDTRKGCCKLSKQDFALAFHLI--NQKLKIDPPHSLT 304
QY 139 PS--EPEDADLTN--SGSSPED-----DALPGSPWRKKLRKKCKEKKKEEPPD 187
DB 305 PEMIPSPDRSSLOKNITGSSPVADPSAIKELDTLNNEIYDLOREKNVNEODLKEKEDTVK 364
QY 188 QDISPLPQPSNRKSRKHTKLEVNKRLODRLSCISPKQH-----SPALQSTDE 242
DB 365 QRTSEV--QDLQDEVQRESINLQKLAQKQVQELLGELDEKQALDEQLQEVRRKCAEE 422
QY 243 VLVV--EGPVLPQSSRLFTLK---IRCRADVLRLPVRMSEPLQNV-----VDHMANHL 290
DB 423 AOLISLKAELITSOEISYEEELKAREELSRLOQETALESVESGKALQLEPQOHL 482
QY 291 GVSPPRIILLFGESELSPTAT-----PSTLKGVDIIDCVLASSSEATETSO 339
DB 483 QESQOETISSMQRLEKDKDETNNOSNMSSSPQSVLVNGATDY--CSLSTSSSETANFNE 540
QY 340 --ELRLVQKKEKHOMLEISLSPD--SPLKVLMSHYEAMGLSGH--KLSEFPDGTLSGKE 395
DB 541 HAEQGNLSEPTHQESSVRSSPEIAPSDV--TDESEAVTAGNEKVTYPRFDDKHSKEE 598
QY 396 LPADLGLESGLIE 409
DB 599 DP--FNVESSSLTD 610

RESULT 11
US-09-061-709-2
Sequence 2, Application US/09061709B
Patent No. 629364
GENERAL INFORMATION:
APPLICANT: Chen, Yao-Tseng
APPLICANT: Gure, All
APPLICANT: Tsang, Solam
APPLICANT: Stockert, Elisabeth
APPLICANT: Jager, Elke
APPLICANT: Knuth, Alexander
TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated
TITLE OF INVENTION: Antigens, The Antigens Per Se, And Uses Thereof
FILE REFERENCE: LUD 5538
CURRENT APPLICATION NUMBER: US/09/061,709B
FILING DATE: 1998-04-17
NUMBER OF SEQ ID NOS: 8

TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864

```

:      TELEX: 66141 PENNIE
:      INFORMATION FOR SEO ID NO: 2:
:      SEQUENCE CHARACTERISTICS:
:      LENGTH: 505 amino acids
:      TYPE: amino acid
:      TOPOLOGY: unknown
:      MOLECULE TYPE: protein
:
US-08-631-200-2

```

Query Match	5.2%;	Score 108.5;	DB 1;	Length 505;
Best Local Similarity	22.9%;	Pred. No. 0.19;		
Matches 64;	Conservative 36;	Mismatches 114;	Indels 65;	Gaps 10;

OY	3	EPLR-----ORGRSRGRRARRARGRCPCAPASPALIPDIYVLVLSDSDEYL	56
Db	44	EPLWQNAADR-PRSR-----RAROSEGA---PLVESLSSSGSTSY	83
OY	57	EVADPEVPVARLPAPAKPEODSDSECGAAE--GPAGAPTLYRRRRRLLDGEARPV	114
Db	84	QVOEADSIASYOLGATRPAPASAKSKSGLAASAOGCAGPERKEKRGHKGSFPA----	138
OY	115	PYSCKVQSSLNLIIPDNS-----SLKLCPSEBDEADILNCSGSPSEDDALPSCSPWR	168
Db	139	-----TLAEKSEAOGPVOILLVGOSDHDKDGETAAGG-----AQPSGDOLR	182
OY	169	KLLRKCKEKEEKMEEPDODISPLPPSSRNKSRSKHTHEALQIKREYNKRLOIDRSCLSP	228
Db	183	ATMQKKGISSMSFDEDED-----ENSSSSQLNSNTRPSSATSRSKSIREAASASP	236
OY	229	KOHOSPALOSTDDEVYLVVEGYVLPOSSKLFITLKIRCRAD	267
Db	237	AAPBP-----VDIEVODLEELARPAPGIGTIKRICTRD	271

```

1 RESULT 1
2
3 ; Sequence 4, Application US/08630592
4 ; Patent No. 5770432
5
6 ; GENERAL INFORMATION:
7
8 ; APPLICANT: Nishina, Patsy
9 ; APPLICANT: No. 5770432entruuth, Konrad
10 ; APPLICANT: Nagert, Juergen
11 ; APPLICANT: No. 5770432tlh, Michael
12 ; TITLE OF INVENTION: Obesity Associated Genes
13 ; NUMBER OF SEQUENCE ADDRESSES: 25
14 ; CORRESPONDENCE ADDRESS:
15 ; ADDRESSEE: FLEHN, HOHACH, TEST, ALBRITTON & HERBERT
16 ; STREET: 3400 Embarcadero Center, Suite 3400
17 ; CITY: San Francisco
18 ; STATE: California
19 ; COUNTRY: USA
20
21 ; ZIP: 94114187
22
23 ; COMPUTER READABLE FORM:
24 ;
25 ; MEDIUM TYPE: Floppy disk
26 ; COMPUTER: IBM PC compatible
27 ; OPERATING SYSTEM: PCDOS/MSDOS
28 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
29
30 ; CURRENT APPLICATION DATA:
31 ; APPLICATION NUMBER: US/08/630,592
32
33 ; FILING DATE:
34
35 ; CLASSIFICATION: 435
36
37 ; ATTORNEY/AGENT INFORMATION:
38 ;
39 ; NAME: Sherwood, Pamela J.
40 ; REGISTRATION NUMBER: 36,677
41 ; REFERENCE/DOCKET NUMBER: A59504/BIR/PJS
42 ; TELECOMMUNICATION INFORMATION:
43 ; TELEPHONE: (415) 7811989
44 ; TELEFAX: (415) 3983249
45 ; TELE: 910 277289
46
47 ; INFORMATION FOR SEQ ID NO: 4:
48
49 ; SEQUENCE CHARACTERISTICS:
50
51 ; LENGTH: 505 amino acids
52
53 ;
54
55 ;
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99

```

```

;      TYPE: amino acid
;      STRANDEDNESS: double
;      TOPOLOGY: linear
;      MOLECULE TYPE: peptide
US-08-630-592-4

```

Query Match	5.2%;	Score 108.5;	DB 1;	Length 505;
Best Local Similarity	22.9%;	Pred. No. 0.19;		
Matches 64;	Conservative 36;	Mismatches 114;	Indels 65;	Gaps 10;

```

0Y 3 EPLR-----GRGRSROGRCARBARGRCSPARPOSPRLPDVLVYVSDSDEVL 56
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 44 EPLWQANADR-PRSR-----RARSQDA---PLVEYSTSSGSTR 83
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y 57 EVADVEVYPARLAPAPAKPEQSDSDSEGAEE--GPAADPTLVRRRRRLPDGEAPV 114
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 84 QVOEADSIASVOLGATRPPAPASAKKSGAASGOGGAPRKEKGNKSTGPA----- 130
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y 115 PYSKQVSSLNLTPDNS-----SLKLCPSPEEDADLTNGSSPSBEDALPSSGPMR 168
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 139 -----TLAEDKSEAQPVOLLTVGSDHDKHAGETPAAGG-----APSGDOLR 182
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y 169 KLLRKCKEKKKEEFPDDQISPLPQPSRNKSRKHTAOLKREYNNKRLDRLSCLP 228
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 183 ATMOKGSISSMSFDEDEDED-----ENSSSSQNLMSNTRPSPGATSRKSIREAASAPSP 236
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
0Y 229 KOHOSPALOSTDDEVYLVLEGVYLPQSSSLFLTKRCRAD 267
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 237 AAPEPP-----VDIEVQDLEEFALPAPGGLTIKCRITRD 271
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

; MOLECULE TYPE: peptide
US-08-714-991-4

Query Match 5.2%; Score 108.5; DB 1; Length 505;
Best Local Similarity 22.9%; Pred. No. 0.19;
Matches 64; Conservative 36; Mismatches 114; Indels 65; Gaps 10;

```

QY 3 EPLR-----GGRSRSGRGARRARGRCRRAROSPAPRLIPDVIYLVDSDEVL 56
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 44 EPLMVQANADGR-PRSR-----RARQSEQA---PLVESYLSSSGSTSY 83

QY 57 EVADPVEVPYARLPAPAKPEQSDSDSEGAEE--GPAGAPRTLVRRRRRLDPGEAPVY 114
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 84 QVQEADSIASVQLGATRRPAPASAKSKGAAAGGOGAPRKEKKGKGTSGPA----- 138

QY 115 PYVSGKVQSSLNLIIPDNS-----SLKLCPSPEDEADLTNGSSPSEDDALPSGSPWR 168
    | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 139 -----TLAEDKSEAQGPVQILTWGSDHDKDAGETAAGG-----AQP SGODLR 182

QY 169 KLLRKKCEKEKEKEEPPDODISPLPOPSSRNKSRKHTEALQKLRVKNRLODLRSGCLSP 228
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 183 ATMQRKGISSMSFDEDEDED-----ENSSSSQLNSNTRPSSATSRKSIREAASAPSP 236

QY 229 KOHQPALOSTDDEVYLVESPVLPQSSRLFTTKIRCRAD 267
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 237 AAPEPE---VDIEVQDLEEFALRPAPQGITIKCRITRD 271

```

Search completed: April 22, 2002, 10:08:37
Job time: 142 sec

This Page Blank (uspio)